STIC-Biotech/ChemLib 171422

From: Swope, Sheridan
Sent: Monday, November 14, 2005 1:20 PM
To: STIC-Biotech/ChemLib
Subject: 09/858,332

For 09/858,332, pls search:

SID 15 against the .rge & .rst databases only
collect the top 300 hits

SID 28 against the .rge & .rst databases only
collect the top 600 hits

Toby Port

Thanks!!

Sheridan Swope, Ph.D.
Patent Examiner, AU 1656
Recombinant Enzymes
571-272-0943 (voice)
E02B71 Remsen Bld (Office)
E03C70 Remsen Bld (Mailbox)

****
Searcher:_____
Searcher Phone:_____
Date Searcher Picked up:_____
Date completed:_____
Searcher Prep Time:_____
Online Time:_____

****
Type of Search
NA#_____ AA#:_____
S/L:_____ Oligomer:_____
Encode/Transl:_____
Structure #:_____Text:____
Inventor:_____ Litigation:___

****
Vendors and cost where applicable
STN:_____
DIALOG:_____
QUESTEL/ORBIT:_____
LEXIS/NEXIS:_____
SEQUENCE SYSTEM:_____
WWW/Internet:_____
Other (Specify):_____

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 28, 2005, 08:17:06 ; Search time 2606 Seconds
(without alignments)
215.443 Million cell updates/sec

Title: US-09-858-332G-15
Perfect score: 84
Sequence: 1 HNHNHNHNHNHN 12

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 600 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09858332/runat_23112005_103443_19703/app_query.fasta_1.199
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=600
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=300 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09858332_@CGN_1_1_5315_@runat_23112005_103443_19703 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	84	100.0	125	3	BJ340611	BJ340611 BJ340611
c 2	84	100.0	174	3	BJ429957	BJ429957 BJ429957
3	84	100.0	209	1	AU053723	AU053723 AU053723
4	84	100.0	214	5	BQ586973	BQ586973 E012352-0
5	84	100.0	264	1	AU053106	AU053106 AU053106
c 6	84	100.0	297	5	BW876703	BW876703 BW876703
7	84	100.0	298	2	BF547803	BF547803 UI-R-E0-b
8	84	100.0	317	8	D47535	D47535 RICS13103A
c 9	84	100.0	318	10	CE715092	CE715092 tigr-gss-
10	84	100.0	319	1	AU061961	AU061961 AU061961
c 11	84	100.0	375	3	BP803155	BP803155 BP803155
12	84	100.0	413	2	BG669342	BG669342 DRNAAED03
13	84	100.0	417	1	AU034405	AU034405 AU034405
c 14	84	100.0	426	1	AI008498	AI008498 EST202949
15	84	100.0	427	1	AU033706	AU033706 AU033706
16	84	100.0	427	1	AU033893	AU033893 AU033893
c 17	84	100.0	446	3	BJ434889	BJ434889 BJ434889
18	84	100.0	453	1	AU040068	AU040068 AU040068
19	84	100.0	459	1	AU036977	AU036977 AU036977
20	84	100.0	466	6	CB714934	CB714934 AMGNNUC:M
c 21	84	100.0	467	3	BJ431306	BJ431306 BJ431306
c 22	84	100.0	472	3	BJ429835	BJ429835 BJ429835
c 23	84	100.0	476	3	BJ431249	BJ431249 BJ431249
c 24	84	100.0	480	11	CR294373	CR294373 mte1-10O1
25	84	100.0	488	9	AZ321550	AZ321550 1M0042I11
26	84	100.0	494	3	BJ388347	BJ388347 BJ388347
27	84	100.0	496	5	C90140	C90140 C90140 Dict
28	84	100.0	500	7	CK041708	CK041708 48024rsic
29	84	100.0	504	7	CK618125	CK618125 ml05h06.y
c 30	84	100.0	528	6	CB264620	CB264620 50-E01465
31	84	100.0	533	9	AQ491972	AQ491972 V100B10 m
32	84	100.0	538	5	C22962	C22962 C22962 Dict
c 33	84	100.0	540	1	AA874794	AA874794 UI-R-E0-b
c 34	84	100.0	546	3	BJ429095	BJ429095 BJ429095
c 35	84	100.0	556	3	BJ430064	BJ430064 BJ430064
c 36	84	100.0	559	2	BG375827	BG375827 UI-R-CS0-
37	84	100.0	560	1	AU265802	AU265802 AU265802
38	84	100.0	560	3	BJ329834	BJ329834 BJ329834
c 39	84	100.0	571	3	BJ430176	BJ430176 BJ430176
c 40	84	100.0	571	3	BJ432285	BJ432285 BJ432285
c 41	84	100.0	577	5	BU886048	BU886048 R039F11 P
c 42	84	100.0	579	6	CB255238	CB255238 46-E01533
43	84	100.0	582	3	BJ369186	BJ369186 BJ369186
44	84	100.0	583	6	CB586337	CB586337 AMGNNUC:N
c 45	84	100.0	592	3	BP796399	BP796399 BP796399
c 46	84	100.0	599	3	BJ429471	BJ429471 BJ429471
c 47	84	100.0	604	3	BJ433687	BJ433687 BJ433687
c 48	84	100.0	606	3	BJ429190	BJ429190 BJ429190
49	84	100.0	609	2	BF237103	BF237103 602027447
c 50	84	100.0	609	3	BJ433986	BJ433986 BJ433986
c 51	84	100.0	609	5	BQ783159	BQ783159 UI-R-FF0-
c 52	84	100.0	611	3	BJ432773	BJ432773 BJ432773
53	84	100.0	616	3	BJ414802	BJ414802 BJ414802
c 54	84	100.0	616	3	BJ428216	BJ428216 BJ428216
c 55	84	100.0	618	3	BJ433790	BJ433790 BJ433790
c 56	84	100.0	620	3	BJ431653	BJ431653 BJ431653
c 57	84	100.0	622	3	BJ401757	BJ401757 BJ401757
c 58	84	100.0	622	3	BJ433898	BJ433898 BJ433898
c 59	84	100.0	624	3	BJ340817	BJ340817 BJ340817
60	84	100.0	625	3	BM778956	BM778956 EST589531
c 61	84	100.0	626	3	BJ428793	BJ428793 BJ428793
c 62	84	100.0	639	3	BJ432833	BJ432833 BJ432833
63	84	100.0	644	11	CR320264	CR320264 mte1-46E2
64	84	100.0	652	2	BI271416	BI271416 NF051D09F
c 65	84	100.0	652	3	BJ432931	BJ432931 BJ432931
c 66	84	100.0	657	3	BJ431681	BJ431681 BJ431681
67	84	100.0	658	6	CD373395	CD373395 UI-R-GR0-
c 68	84	100.0	661	5	BQ781099	BQ781099 UI-R-FF0-
c 69	84	100.0	662	2	BI281721	BI281721 UI-R-CT0s
c 70	84	100.0	663	1	AV823854	AV823854 AV823854
c 71	84	100.0	663	3	BJ429328	BJ429328 BJ429328
c 72	84	100.0	667	3	BJ433069	BJ433069 BJ433069
c 73	84	100.0	668	3	BJ430907	BJ430907 BJ430907
c 74	84	100.0	668	3	BJ434653	BJ434653 BJ434653
c 75	84	100.0	676	3	BJ431437	BJ431437 BJ431437
76	84	100.0	676	7	CV521375	CV521375 0089P0058
c 77	84	100.0	677	3	BJ429677	BJ429677 BJ429677
78	84	100.0	678	2	BG448016	BG448016 NF104H02E
c 79	84	100.0	679	3	BJ431232	BJ431232 BJ431232
c 80	84	100.0	690	3	BJ430517	BJ430517 BJ430517

	81	84	100.0	695	10	CG931836
c	82	84	100.0	697	3	BJ434010
	83	84	100.0	697	7	CO389214
c	84	84	100.0	701	3	BJ432382
c	85	84	100.0	714	9	AQ874103
c	86	84	100.0	717	3	BJ430955
c	87	84	100.0	718	3	BJ445670
c	88	84	100.0	723	3	BJ429204
c	89	84	100.0	724	3	BJ434636
c	90	84	100.0	725	3	BJ429573
	91	84	100.0	725	7	CN834352
c	92	84	100.0	731	3	BJ436240
	93	84	100.0	735	11	CR277961
	94	84	100.0	736	2	BF385371
	95	84	100.0	738	5	C93043
	96	84	100.0	738	7	CO387676
c	97	84	100.0	741	3	BJ401045
c	98	84	100.0	749	10	BX211625
c	99	84	100.0	752	10	CZ281894
c	100	84	100.0	759	3	BJ434227
	101	84	100.0	760	10	AG542244
	102	84	100.0	761	7	CO427845
c	103	84	100.0	762	3	BJ430651
c	104	84	100.0	763	3	BJ431956
c	105	84	100.0	768	3	BJ432448
c	106	84	100.0	774	3	BJ431981
	107	84	100.0	774	7	CO426813
	108	84	100.0	775	7	CO393602
c	109	84	100.0	778	10	AG359323
	110	84	100.0	783	7	CK480095
	111	84	100.0	797	7	CK596357
	112	84	100.0	802	7	CO560704
	113	84	100.0	811	6	CB313044
	114	84	100.0	811	8	DN475931
	115	84	100.0	832	7	CK469853
	116	84	100.0	836	7	CO396373
	117	84	100.0	854	7	CK470094
	118	84	100.0	898	2	BF384538
	119	84	100.0	913	2	BI247099
	120	84	100.0	924	2	BI147575
	121	84	100.0	929	5	BQ946631
c	122	84	100.0	938	10	AG403889
	123	84	100.0	941	2	BI145320
	124	84	100.0	963	2	BF384348
	125	84	100.0	969	10	CNS01S89
c	126	84	100.0	983	4	CNS0A3V5
	127	84	100.0	1127	10	CW926142
c	128	84	100.0	1147	10	CG748246
	129	84	100.0	1260	10	CG751785
	130	79	94.0	228	7	CO484355
c	131	79	94.0	357	6	CD852815
	132	79	94.0	358	3	BM300074
c	133	79	94.0	363	9	BH142412
c	134	79	94.0	427	3	BP805483
c	135	79	94.0	510	6	CA890073
c	136	79	94.0	532	7	CK844809
c	137	79	94.0	562	9	BH142693
c	138	79	94.0	635	10	CZ285358
	139	79	94.0	645	1	AU268952
	140	79	94.0	649	8	CX535222
c	141	79	94.0	674	10	CZ284612
	142	79	94.0	743	10	CZ288826
c	143	79	94.0	755	3	BJ434857
c	144	79	94.0	756	10	CZ288827
c	145	79	94.0	783	10	CZ288255
	146	79	94.0	970	9	CC653644
	147	79	94.0	1027	11	CNS06LYT
	148	79	94.0	1126	11	CNS06LAO
c	149	78	92.9	253	10	CG895908
c	150	78	92.9	338	10	CG898233
c	151	78	92.9	349	7	CV252723
c	152	78	92.9	376	3	BP866389
c	153	78	92.9	430	5	BU017925

CG931836 MBEAL86TR
BJ434010 BJ434010
CO389214 AGENCOURT
BJ432382 BJ432382
AQ874103 V103A5 mT
BJ430955 BJ430955
BJ445670 BJ445670
BJ429204 BJ429204
BJ434636 BJ434636
BJ429573 BJ429573
CN834352 AGENCOURT
BJ436240 BJ436240
CR277961 Forward s
BF385371 602047107
C93043 C93043 Dict
CO387676 AGENCOURT
BJ401045 BJ401045
BX211625 Danio rer
CZ281894 cp19e05.f
BJ434227 BJ434227
AG542244 Mus muscu
CO427845 UI-M-HW0-
BJ430651 BJ430651
BJ431956 BJ431956
BJ432448 BJ432448
BJ431981 BJ431981
CO426813 UI-M-HU0-
CO393602 AGENCOURT
AG359323 Mus muscu
CK480095 AGENCOURT
CK596357 AGENCOURT
CO560704 AGENCOURT
CB313044 AGENCOURT
DN475931 altr202xg
CK469853 AGENCOURT
CO396373 AGENCOURT
CK470094 AGENCOURT
BF384538 602046807
BI247099 602960664
BI147575 602913017
BQ946631 AGENCOURT
AG403889 Mus muscu
BI145320 602910458
BF384348 602046701
AL164898 Tetraodon
BX828453 Arabidops
CW926142 EDCBE62TF
CG748246 P042-1-D0
CG751785 P046-2-C0
CO484355 GQ0204.B3
CD852815 DH0ALL9ZH
BM300074 MCR054H05
BH142412 TDGDO50TH
BP805483 BP805483
CA890073 B0156H06-
CK844809 UI-R-AE0-
BH142693 TDGDR64TH
CZ285358 cp40g07.f
AU268952 AU268952
CX535222 s13dNF95H
CZ284612 cp36e12.r
CZ288826 cp60d09.f
BJ434857 BJ434857
CZ288827 cp60d09.r
CZ288255 cp57c02.r
CC653644 OGUIJ89TV
AL404811 T7 end of
AL403942 T7 end of
CG895908 pastbac00
CG898233 pastbac06
CV252723 PX0015.B2
BP866389 BP866389
BU017925 QHE17D20.

	154	78	92.9	511	11	CR297245
	155	78	92.9	543	1	AW736366
	156	78	92.9	573	7	CN884225
	157	78	92.9	663	3	BJ325871
	158	78	92.9	691	6	CB676683
	159	78	92.9	694	7	CO363639
	160	78	92.9	795	8	DR693890
	161	78	92.9	851	8	DR631262
	162	78	92.9	1125	2	BG298931
	163	77	91.7	250	6	CB606620
c	164	77	91.7	314	3	BJ351548
c	165	77	91.7	327	3	BJ437953
	166	77	91.7	519	2	BG362662
c	167	77	91.7	520	6	CA899924
	168	77	91.7	531	1	AU053519
c	169	77	91.7	717	3	BJ384768
	170	76	90.5	247	5	BU994168
	171	76	90.5	334	5	BQ470935
c	172	76	90.5	390	1	AW332717
	173	76	90.5	516	5	CA023404
	174	76	90.5	527	2	BF254321
	175	76	90.5	562	5	BU494816
	176	76	90.5	566	5	BU997619
	177	76	90.5	587	6	CF638492
	178	76	90.5	597	6	CA855800
	179	76	90.5	600	6	CA854726
	180	76	90.5	612	6	CA837970
	181	76	90.5	671	6	CA855323
	182	76	90.5	697	2	BE060493
	183	76	90.5	727	5	BU561234
	184	76	90.5	745	6	CF638604
	185	76	90.5	748	2	BF235379
	186	75	89.3	322	5	BU275786
c	187	75	89.3	487	10	CNS026FK
	188	75	89.3	655	10	CW461432
	189	75	89.3	662	10	CW233590
c	190	75	89.3	784	10	CL405652
	191	74	88.1	291	10	CL960835
	192	74	88.1	368	7	CF963928
	193	74	88.1	386	2	BF068492
	194	74	88.1	395	6	CF196074
	195	74	88.1	405	6	CF195899
c	196	74	88.1	420	9	AQ655474
	197	74	88.1	426	3	BI798061
	198	74	88.1	430	7	CF987483
	199	74	88.1	437	7	CF965114
	200	74	88.1	438	7	CF986640
	201	74	88.1	438	7	CK080515
	202	74	88.1	440	1	AW596891
	203	74	88.1	448	7	CF965703
	204	74	88.1	449	7	CF960519
c	205	74	88.1	449	7	CK746439
	206	74	88.1	460	6	CF197290
	207	74	88.1	471	7	CF985613
	208	74	88.1	475	7	CF962395
	209	74	88.1	476	7	CF961968
c	210	74	88.1	476	9	AQ659741
	211	74	88.1	486	7	CF961117
	212	74	88.1	491	7	CF963249
c	213	74	88.1	492	6	CB964758
	214	74	88.1	495	7	CF960382
	215	74	88.1	497	7	CF988679
	216	74	88.1	499	7	CF991700
	217	74	88.1	505	7	CF963876
	218	74	88.1	514	7	CF965313
	219	74	88.1	516	7	CF985124
c	220	74	88.1	518	5	BU673005
c	221	74	88.1	518	5	BU673252
c	222	74	88.1	518	6	CB965151
	223	74	88.1	521	7	CF964678
	224	74	88.1	529	7	CF985622
c	225	74	88.1	531	6	CB965275
c	226	74	88.1	531	6	CB967418

CR297245 mtel-14A4
AW736366 EST332285
CN884225 010815AAS
BJ325871 BJ325871
CB676683 OSJNBe13B
CO363639 RTK1_10_H
DR693890 EST108398
DR631262 EST102139
BG298931 602397468
CB606620 AMGNNUC:C
BJ351548 BJ351548
BJ437953 BJ437953
BG362662 sac07a11.
CA899924 PCS00167M
AU053519 AU053519
BJ384768 BJ384768
BU994168 HM05P14r
BQ470935 HX04H16r
AW332717 S12A5 AGS
CA023404 HZ46C09r
BF254321 HVSMEf000
BU494816 PfESToab6
BU997619 HI08I06r
CF638492 D03_C02 F
CA855800 PfESToac3
CA854726 PfESToac1
CA837970 MCT011D06
CA855323 PfESToac3
BE060493 HVSMEg001
BU561234 AGENCOURT
CF638604 D04_F10 F
BF235379 602026773
BU275786 Cr_Emb_07
AL183305 Tetraodon
CW461432 fsbb001f2
CW233590 104_687_1
CL405652 ZMMBBb039
CL960835 OsIFCC037
CF963928 13692rsic
BF068492 st82f08.y
CF196074 Azbi2.3_E
CF195899 Azbi2.1_A
AQ655474 Sheared_D
BI798061 H096C06 E
CF987483 15498rsic
CF965114 9492rsice
CF986640 23319rsic
CK080515 81076rsic
AW596891 sj84b04.y
CF965703 14418rsic
CF960519 12399rsic
CK746439 wmi01-8ms
CF197290 Ir64vi2.1
CF985613 17031rsic
CF962395 9808siced
CF961968 13885rsic
AQ659741 Sheared D
CF961117 11158rsic
CF963249 12143rsic
CB964758 NL11_D05
CF960382 12469rsic
CF988679 17338rsic
CF991700 16342rsic
CF963876 14637rsic
CF965313 11490rsic
CF985124 20265rsic
BU673005 NL_9_53 D
BU673252 NL_13_3 D
CB965151 NL17_E02
CF964678 12289rsic
CF985622 17719rsic
CB965275 NL19_A06
CB967418 NL52_B12

c	227	74	88.1	531	6	CB967427	CB967427 NL52_D11
	228	74	88.1	532	3	BI798956	BI798956 H121D04 E
	229	74	88.1	537	7	CK082135	CK082135 80962rsic
	230	74	88.1	542	7	CF960953	CF960953 8110rsice
	231	74	88.1	542	7	CF966553	CF966553 10790rsic
	232	74	88.1	543	7	CF966531	CF966531 7915rsice
	233	74	88.1	543	7	CF991407	CF991407 18863rsic
	234	74	88.1	546	7	CF965538	CF965538 7941rsice
	235	74	88.1	548	7	CF961410	CF961410 12411rsic
c	236	74	88.1	549	5	BX900515	BX900515
c	237	74	88.1	549	6	CB967111	CB967111 NL48_B05
	238	74	88.1	549	7	CF961941	CF961941 13585rsic
c	239	74	88.1	550	5	BU673593	BU673593 NL_17_56
c	240	74	88.1	550	5	BU673632	BU673632 NL_18_10
c	241	74	88.1	550	6	CB964701	CB964701 NL_10C08
c	242	74	88.1	550	6	CB966849	CB966849 NL43_F12
c	243	74	88.1	555	7	CK758412	CK758412 wmi01-10m
c	244	74	88.1	560	7	CK766845	CK766845 wmi01-6ms
	245	74	88.1	562	7	CF963651	CF963651 8463rsice
	246	74	88.1	563	7	CF966423	CF966423 9528rsice
	247	74	88.1	564	7	CF961387	CF961387 13695rsic
	248	74	88.1	565	3	BM143854	BM143854 saj49e11.
	249	74	88.1	565	7	CF959693	CF959693 7904rsice
	250	74	88.1	565	7	CF965105	CF965105 13519rsic
	251	74	88.1	567	6	CA486806	CA486806 WHE4336_B
c	252	74	88.1	570	6	CD491431	CD491431 WHE3087_E
	253	74	88.1	578	2	BE195039	BE195039 HVSMEh008
	254	74	88.1	579	7	CF959561	CF959561 9814siced
	255	74	88.1	583	7	CF984909	CF984909 18592sice
	256	74	88.1	583	7	CF991926	CF991926 16250rsic
	257	74	88.1	584	7	CK082860	CK082860 85761rsic
	258	74	88.1	588	7	CF967065	CF967065 9827siced
	259	74	88.1	593	7	CF953321	CF953321 6565rsice
	260	74	88.1	594	7	CK079978	CK079978 79610rsic
	261	74	88.1	601	10	CZ563660	CZ563660 1_p_bW038
	262	74	88.1	604	3	BJ397270	BJ397270 BJ397270
	263	74	88.1	610	7	CK080511	CK080511 84855rsic
	264	74	88.1	611	7	CV579577	CV579577 kc89h11.y
	265	74	88.1	617	7	CF963590	CF963590 9842siced
	266	74	88.1	643	2	BF108353	BF108353 EI31K10 N
c	267	74	88.1	715	10	CZ292359	CZ292359 cp80h11.f
	268	74	88.1	792	5	BX900547	BX900547 BX900547
c	269	74	88.1	798	8	DR027045	DR027045 Osmo01078
	270	74	88.1	803	7	CK251091	CK251091 EST734728
	271	74	88.1	804	8	CX114190	CX114190 EI086F05
c	272	74	88.1	813	8	CX669339	CX669339 UCRCP01_0
	273	74	88.1	818	8	CX118645	CX118645 EI116B05
c	274	74	88.1	854	10	CZ287109	CZ287109 cp50f12.f
c	275	74	88.1	862	10	CZ292089	CZ292089 cp79d09.r
	276	74	88.1	876	8	CX669338	CX669338 UCRCP01_0
	277	74	88.1	899	8	CX104704	CX104704 BI104A04
	278	74	88.1	899	8	CX112129	CX112129 EI074E01
	279	74	88.1	932	10	DU040292	DU040292 25505 Tom
	280	74	88.1	942	8	CX118565	CX118565 EI115L11
	281	74	88.1	989	11	CNS07CAB	AL438921 T3 end of
c	282	74	88.1	1032	10	AG612034	AG612034 Mus muscu
	283	74	88.1	1076	8	CX115929	CX115929 EI101J02
	284	74	88.1	1171	2	BE035458	BE035458 MO06B10 M
	285	74	88.1	1314	8	CX109102	CX109102 RECm0303
	286	73.5	87.5	395	1	AU040007	AU040007 AU040007
	287	73.5	87.5	718	1	AU034181	AU034181 AU034181
	288	73	86.9	290	7	CF986668	CF986668 24212rsic
	289	73	86.9	315	7	CK116691	CK116691 B015P72 H
	290	73	86.9	316	8	H35592	H35592 EST110865 R
	291	73	86.9	372	7	CF991958	CF991958 17088rsic
	292	73	86.9	397	7	CF966741	CF966741 12584rsic
	293	73	86.9	416	1	AV818700	AV818700 AV818700
	294	73	86.9	419	6	CF096176	CF096176 QHN22F11.
	295	73	86.9	449	3	BP619865	BP619865 BP619865
	296	73	86.9	457	7	CF985050	CF985050 17245rsic
c	297	73	86.9	478	6	CB257174	CB257174 40-E01274
c	298	73	86.9	525	11	CR888025	CR888025 Sus scrof
c	299	73	86.9	591	6	CB257064	CB257064 47-E01274
	300	73	86.9	667	1	AJ803321	AJ803321 AJ803321
c	301	73	86.9	669	3	BJ348434	BJ348434 BJ348434
c	302	73	86.9	697	9	B26426	B26426 F2I19TF IGF
	303	73	86.9	728	8	DT104655	DT104655 JGI_ANNN7
c	304	73	86.9	746	3	BJ376527	BJ376527 BJ376527
c	305	73	86.9	841	4	CNS09Z7X	BX832854 Arabidops
c	306	73	86.9	1020	10	CNS01ELE	AL140691 Anopheles
c	307	72.5	86.3	927	11	CNS06L0N	AL403581 T3 end of
c	308	72.5	86.3	927	11	CNS06L0R	AL403585 T3 end of
	309	72	85.7	348	6	CF196076	CF196076 Azbi2.3_E
	310	72	85.7	427	8	DR699194	DR699194 An_3235.A
	311	72	85.7	438	2	BE758972	BE758972 an_3235 A
	312	72	85.7	544	1	AW638381	AW638381 bl69d02.w
c	313	72	85.7	712	7	CK955000	CK955000 4094901 B
	314	72	85.7	723	11	CR083504	CR083504 Forward s
c	315	72	85.7	738	9	BZ203454	BZ203454 CH230-399
c	316	72	85.7	738	10	CZ281246	CZ281246 cp15h07.f
	317	72	85.7	804	10	CZ282928	CZ282928 cp27a11.r
	318	72	85.7	903	2	BI246731	BI246731 602959649
c	319	72	85.7	920	10	CG957699	CG957699 MBEMD68TR
	320	72	85.7	935	10	CNS026U7	AL183832 Tetraodon
	321	72	85.7	943	7	CK411318	CK411318 AUF_IpHdk
	322	72	85.7	944	7	CK416387	CK416387 AUF_IpInt
	323	72	85.7	994	7	CK405269	CK405269 AUF_IfSpn
	324	71	84.5	324	1	AW664206	AW664206 hi06c06.x
c	325	71	84.5	431	2	BG071356	BG071356 H3097H01-
c	326	71	84.5	436	3	BM369212	BM369212 EBem07_SQ
c	327	71	84.5	457	3	BP579457	BP579457 BP579457
c	328	71	84.5	461	5	BQ853517	BQ853517 QGB20L16.
	329	71	84.5	499	7	CV548502	CV548502 3032HAEME
	330	71	84.5	520	6	CF227970	CF227970 PtaXM0007
c	331	71	84.5	532	1	AL045178	AL045178 DKFZp434B
c	332	71	84.5	545	9	AQ385558	AQ385558 RPCI11-14
	333	71	84.5	549	6	CB541711	CB541711 PVEPSE300
	334	71	84.5	593	9	AZ452610	AZ452610 1M0252L04
	335	71	84.5	665	7	CV539579	CV539579 POD_020_G
	336	71	84.5	670	2	BG084224	BG084224 H3097H01-
	337	71	84.5	686	9	AZ452598	AZ452598 1M0252J04
	338	71	84.5	689	5	BU005046	BU005046 QGG6P01.y
	339	71	84.5	692	7	CV539538	CV539538 POD_020_C
	340	71	84.5	721	2	BG080322	BG080322 H3052C06-
c	341	71	84.5	721	7	CO106749	CO106749 GR__Eb003
	342	71	84.5	725	5	BQ867022	BQ867022 QGC9J06.y
	343	71	84.5	731	7	CV534723	CV534723 NOD_207_G
c	344	71	84.5	753	9	BZ467746	BZ467746 BONOO67TR
c	345	71	84.5	775	9	BZ199802	BZ199802 CH230-303
c	346	71	84.5	783	10	CZ290750	CZ290750 cp71g03.r
c	347	71	84.5	794	10	CZ955988	CZ955988 272597 To
c	348	71	84.5	800	10	CZ280894	CZ280894 cp13h12.r
c	349	71	84.5	804	10	CZ287750	CZ287750 cp54d04.f
	350	71	84.5	816	8	DR940860	DR940860 EST113239
	351	71	84.5	819	2	BF629530	BF629530 HVSMEb001
c	352	71	84.5	824	10	CZ283027	CZ283027 cp27f03.f
c	353	71	84.5	824	10	CZ289925	CZ289925 cp67a03.r
c	354	71	84.5	838	10	CZ289863	CZ289863 cp66f04.r
c	355	71	84.5	852	8	DR944556	DR944556 EST113609
c	356	71	84.5	907	8	DR927540	DR927540 EST111907
	357	71	84.5	924	10	CZ953727	CZ953727 265974 To
	358	71	84.5	943	10	CZ366184	CZ366184 ZMMBF0146
c	359	71	84.5	1360	10	AG448342	AG448342 Mus muscu
	360	70	83.3	292	1	AJ645412	AJ645412 AJ645412
	361	70	83.3	322	10	CE737859	CE737859 tigr-gss-
	362	70	83.3	340	10	CZ490360	CZ490360 f07622-3p
c	363	70	83.3	373	2	BG662150	BG662150 Ljirnpest
	364	70	83.3	415	7	CO543146	CO543146 LyEST1047
c	365	70	83.3	476	3	BP529011	BP529011 BP529011
	366	70	83.3	491	3	BP043778	BP043778 BP043778
	367	70	83.3	499	5	BQ489630	BQ489630 21-E9233-
c	368	70	83.3	516	1	AI600077	AI600077 EST251780
c	369	70	83.3	585	10	AG945847	AG945847 Drosophil
	370	70	83.3	593	6	CA835626	CA835626 MCS039B10
	371	70	83.3	596	1	AU038027	AU038027 AU038027
	372	70	83.3	604	3	BI421372	BI421372 EST532038

	373	70	83.3	625	5	BQ586753
	374	70	83.3	632	3	BJ361063
	375	70	83.3	644	3	BJ327539
	376	70	83.3	651	2	BI263263
	377	70	83.3	676	6	CA838378
c	378	70	83.3	688	6	CA483196
	379	70	83.3	690	11	CNS049P2
	380	70	83.3	691	6	CA840142
	381	70	83.3	713	3	BI930662
c	382	70	83.3	737	7	CV534903
c	383	70	83.3	744	9	AQ854783
c	384	70	83.3	761	9	AZ905920
c	385	70	83.3	793	9	CC974069
c	386	70	83.3	885	7	CV288171
	387	70	83.3	911	10	DU052854
	388	70	83.3	1137	10	CW925680
c	389	69	82.1	234	7	CK132391
	390	69	82.1	265	1	AU061065
	391	69	82.1	307	2	BG046168
c	392	69	82.1	374	3	BI306132
c	393	69	82.1	391	3	BP848938
	394	69	82.1	393	9	AZ254386
	395	69	82.1	405	2	BF324706
	396	69	82.1	416	3	BM270929
	397	69	82.1	420	7	CF991917
	398	69	82.1	425	5	BQ576463
c	399	69	82.1	427	1	AV596087
c	400	69	82.1	438	3	BP583463
	401	69	82.1	442	3	BM577130
c	402	69	82.1	458	1	AW332190
c	403	69	82.1	467	5	BU545370
c	404	69	82.1	473	9	BH357772
	405	69	82.1	514	3	BM886283
	406	69	82.1	515	5	BU780134
	407	69	82.1	539	9	AQ690487
	408	69	82.1	540	7	CF961319
	409	69	82.1	554	7	CF962808
	410	69	82.1	554	8	DN648455
	411	69	82.1	554	9	AZ145694
	412	69	82.1	558	7	CF965219
	413	69	82.1	565	5	BW826927
	414	69	82.1	572	6	CF311500
	415	69	82.1	584	5	BQ755644
c	416	69	82.1	588	1	AW334163
	417	69	82.1	590	3	BI636340
c	418	69	82.1	593	1	AW334680
	419	69	82.1	604	3	BM591087
	420	69	82.1	614	6	CD476478
	421	69	82.1	615	3	BJ421626
	422	69	82.1	627	6	CF318123
	423	69	82.1	633	3	BM619679
	424	69	82.1	637	3	BM637449
	425	69	82.1	638	3	BM651668
	426	69	82.1	642	3	BJ328920
	427	69	82.1	642	7	CV161943
	428	69	82.1	645	7	CV161987
	429	69	82.1	648	3	BM598375
	430	69	82.1	652	7	CK658795
	431	69	82.1	653	3	BM607311
	432	69	82.1	660	7	CV303352
c	433	69	82.1	665	1	AW335593
	434	69	82.1	665	3	BM656053
	435	69	82.1	669	7	CV527620
c	436	69	82.1	670	3	BP130435
c	437	69	82.1	672	9	AZ523539
	438	69	82.1	674	7	CV303516
	439	69	82.1	676	2	BE661545
	440	69	82.1	676	10	CW257991
c	441	69	82.1	686	7	CK132074
	442	69	82.1	690	6	CF307672
	443	69	82.1	692	3	BM577422
	444	69	82.1	695	5	BQ875251
	445	69	82.1	696	7	CV479322

BQ586753 E012389-0
BJ361063 BJ361063
BJ327539 BJ327539
BI263263 NF087C11P
CA838378 MCT015H04
CA483196 LuP12007A
AL280847 Tetraodon
CA840142 MCT040C01
BI930662 EST550551
CV534903 NOD_210_B
AQ854783 CpG2155B
AZ905920 RPCI-24-1
CC974069 ZUAF443TV
CV288171 aof01-5ms
DU052854 140243 To
CW925680 EDCBC04TF
CK132391 LP13887.5
AU061065 AU061065
BG046168 saa49e08.
BI306132 NL_3_H06
BP848938 BP848938
AZ254386 Gm_ISb001
BF324706 sul4c11.y
BM270929 sak03c05.
CF991917 23707rsic
BQ576463 sap17e08.
AV596087 AV596087
BP583463 BP583463
BM577130 170006871
AW332190 S5D3 AGS-
BU545370 GM880004B
BH357772 CH230-95B
BM886283 sam14f07.
BU780134 SJEFUA07
AQ690487 nbxb0082K
CF961319 10864rsic
CF962808 12292rsic
DN648455 G6785.68
AZ145694 SP_0025_A
CF965219 8104rsice
BW826927 BW826927
CF311500 ABF--06-L
BQ755644 EBed07_SQ
AW334163 S31C9 AGS
BI636340 SD17985.5
AW334680 S39B3 AGS
BM591087 170006873
CD476478 eca01-32m
BJ421626 BJ421626
CF318123 HD--08-B1
BM619679 170006874
BM637449 170006875
BM651668 170006873
BJ328920 BJ328920
CV161943 CS_gil_49
CV161987 CS_gil_49
BM598375 170006876
CK658795 LP16040.5
BM607311 170006870
CV303352 CS_hyp_32
AW335593 S49B9 AGS
BM656053 170006873
CV527620 CS_GIL_06
BP130435 BP130435
AZ523539 220PbG03
CV303516 CS_hyp_32
BE661545 2-F9 Gmax
CW257991 104_723_1
CK132074 LP13314.5
CF307672 ABF--01-C
BM577422 170006871
BQ875251 QGI7J09.y
CV479322 CS_hyp_43

	446	69	82.1	709	3	BM602325
	447	69	82.1	710	3	BM638080
	448	69	82.1	718	7	CV527141
c	449	69	82.1	737	9	BZ093668
	450	69	82.1	846	9	AQ500671
c	451	69	82.1	911	10	CW919353
	452	69	82.1	937	10	CNS02L50
	453	69	82.1	959	4	AY810820
	454	68	81.0	415	7	CF967055
	455	68	81.0	447	6	CA410761
	456	68	81.0	450	7	CF963362
	457	68	81.0	455	7	CF989857
	458	68	81.0	463	7	CF992233
	459	68	81.0	481	10	CZ741783
	460	68	81.0	498	7	CF966833
c	461	68	81.0	501	3	BJ434315
	462	68	81.0	502	7	CF990259
	463	68	81.0	519	7	CF986055
	464	68	81.0	528	7	CF963854
	465	68	81.0	529	7	CK035650
	466	68	81.0	544	6	CB967041
	467	68	81.0	561	7	CF964295
	468	68	81.0	563	7	CF988027
c	469	68	81.0	569	10	CG289925
	470	68	81.0	573	7	CF990911
	471	68	81.0	575	7	CF988960
	472	68	81.0	578	7	CF988044
	473	68	81.0	587	7	CK081259
	474	68	81.0	587	8	CV850474
	475	68	81.0	590	7	CF965643
	476	68	81.0	600	5	BQ997973
	477	68	81.0	639	10	CG436003
	478	68	81.0	651	2	BF142043
	479	68	81.0	659	10	CZ379685
	480	68	81.0	682	9	CC983181
c	481	68	81.0	699	8	CX292217
	482	68	81.0	744	10	CG401758
c	483	68	81.0	750	10	CZ289191
	484	68	81.0	810	7	CN593275
c	485	68	81.0	815	11	CNS07DF2
c	486	68	81.0	837	7	CK936217
c	487	68	81.0	840	10	CNS02X4Z
	488	68	81.0	844	7	CN595432
	489	68	81.0	855	10	CG289909
	490	68	81.0	857	7	CN596342
	491	68	81.0	876	8	DR948273
c	492	68	81.0	883	10	CZ326741
	493	68	81.0	905	7	CF885120
	494	68	81.0	965	10	CZ347441
c	495	68	81.0	978	9	CC975417
	496	68	81.0	996	5	BY703077
	497	68	81.0	1039	11	CNS07DH0
	498	68	81.0	1042	8	CX114449
	499	68	81.0	1927	4	AK002562
	500	67.5	80.4	640	9	AZ195910
	501	67.5	80.4	681	7	CV521814
c	502	67.5	80.4	712	8	DT046127
c	503	67.5	80.4	761	8	DT046193
c	504	67	79.8	125	7	CO516793
	505	67	79.8	234	2	BF428729
	506	67	79.8	346	7	CO267388
c	507	67	79.8	361	3	BJ340157
c	508	67	79.8	368	1	AJ761032
c	509	67	79.8	383	2	BG588299
	510	67	79.8	396	7	CF959995
c	511	67	79.8	446	1	AL379537
c	512	67	79.8	508	1	AL368573
	513	67	79.8	537	1	AJ610887
	514	67	79.8	542	2	BE429782
	515	67	79.8	549	3	BJ560243
c	516	67	79.8	553	1	AL368060
c	517	67	79.8	581	7	CK660246
	518	67	79.8	624	1	AW586237

BM602325 170006870
BM638080 170006875
CV527141 CS_GIL_10
BZ093668 CH230-233
AQ500671 V34H11 mT
CW919353 EDCAA38TF
AL202365 Tetraodon
AY810820 Schistoso
CF967055 7312rsice
CA410761 E031_F -P
CF963362 13642rsic
CF989857 24329rsic
CF992233 16866rsic
CZ741783 OC__Ba008
CF966833 9382rsice
BJ434315 BJ434315
CF990259 15651rsic
CF986055 22359rsic
CF963854 12989rsic
CK035650 37681rsic
CB967041 NL47_A04
CF964295 8010rsice
CF988027 19150rsic
CG289925 OGWLU09TV
CF990911 19801rsic
CF988960 20647rsic
CF988044 24006rsic
CK081259 80406rsic
CV850474 zaa73d11.
CF965643 10188rsic
BQ997973 QGG17P07.
CG436003 OGTBX18TV
BF142043 601790643
CZ379685 ZMMBF0152
CC983181 ZUAC012TV
CX292217 C04011H03
CG401758 ZMMBBb024
CZ289191 cp62f06.r
CN593275 TTE000102
AL440388 T3 end of
CK936217 CGF100438
AL217916 Tetraodon
CN595432 TTE000134
CG289909 OGWLU09TH
CN596342 TTE000072
DR948273 EST113981
CZ326741 ZMMBF0032
CF885120 tric077xn
CZ347441 ZMMBF0121
CC975417 ZUAA005TH
BY703077 BY703077
AL440458 T3 end of
CX114449 EI088A06
AK002562 Mus muscu
AZ195910 SP_1031_A
CV521814 0089P0065
DT046127 Le_lv0_38
DT046193 Le_lv0_39
CO516793 s13dSG94E
BF428729 WHE1412_F
CO267388 EK150015.
BJ340157 BJ340157
AJ761032 AJ761032
BG588299 EST490108
CF959995 8512rsice
AL379537 MtBB45H08
AL368573 MtBA25C11
AJ610887 AJ610887
BE429782 TAS003.G0
BJ560243 BJ560243
AL368060 MtBA21H11
CK660246 LP18649.5
AW586237 EST317860

	519	67	79.8	649	2	BE426346 BE426346 WHE0334_G
	520	67	79.8	701	7	CN244424 CN244424 EST010301
	521	67	79.8	819	8	CX589481 CX589481 TTE000219
	522	67	79.8	858	9	CC599593 CC599593 ZMMBBc040
c	523	67	79.8	860	10	CL009963 CL009963 ZMMBBb055
c	524	67	79.8	995	11	CNS070WM AL424172 T7 end of
c	525	67	79.8	1069	7	CK214039 CK214039 FGAS02595
c	526	66.5	79.2	301	6	CB700752 CB700752 AMGNNUC:S
	527	66.5	79.2	717	8	DR932536 DR932536 EST112407
	528	66	78.6	161	6	CB214559 CB214559 OML04839
	529	66	78.6	220	6	CB214364 CB214364 OML04644
c	530	66	78.6	287	7	CO813747 CO813747 AGENCOURT
	531	66	78.6	336	6	CB211509 CB211509 OML01789
	532	66	78.6	337	5	C93386 C93386 C93386 Dict
	533	66	78.6	337	6	CB212482 CB212482 OML02762
	534	66	78.6	362	7	CF990498 CF990498 22837rsic
c	535	66	78.6	363	3	BI305814 BI305814 NL_1_N10
c	536	66	78.6	364	3	BI306083 BI306083 NL_3_E04
c	537	66	78.6	365	3	BI306516 BI306516 NL_5_A01
c	538	66	78.6	432	10	CW855196 CW855196 shS2kd11-
c	539	66	78.6	437	6	CB964579 CB964579 NL8_E02 D
	540	66	78.6	442	7	CF988556 CF988556 18328rsic
	541	66	78.6	445	7	CF961353 CF961353 7703rsice
	542	66	78.6	446	3	BJ324275 BJ324275 BJ324275
	543	66	78.6	509	7	CF991173 CF991173 18042rsic
	544	66	78.6	529	3	BJ324484 BJ324484 BJ324484
	545	66	78.6	557	6	CB210072 CB210072 OML00352
	546	66	78.6	558	7	CF988053 CF988053 15729rsic
	547	66	78.6	567	3	BM519961 BM519961 sak86g10.
c	548	66	78.6	579	3	BJ445045 BJ445045 BJ445045
	549	66	78.6	604	7	CO334768 CO334768 EK315242.
	550	66	78.6	620	8	CX044463 CX044463 UCRCS07_1
c	551	66	78.6	645	8	CV996341 CV996341 IpCGFr4_1
c	552	66	78.6	647	8	DR660731 DR660731 EST105084
	553	66	78.6	652	7	CO131250 CO131250 GR__Eb43H
	554	66	78.6	654	7	CV717767 CV717767 UCRCS08_0
	555	66	78.6	674	11	CR035801 CR035801 Reverse s
c	556	66	78.6	685	7	CN787661 CN787661 4121846 B
c	557	66	78.6	700	3	BJ407102 BJ407102 BJ407102
c	558	66	78.6	702	3	BJ443417 BJ443417 BJ443417
	559	66	78.6	703	6	CD038320 CD038320 UTPPI005_
	560	66	78.6	738	9	CC828437 CC828437 ZMMBBb017
c	561	66	78.6	741	10	CL689743 CL689743 PRI0152a_
	562	66	78.6	784	8	DR041357 DR041357 BNE000013
c	563	66	78.6	803	10	CL676387 CL676387 PRI0118b_
c	564	66	78.6	814	6	CF269740 CF269740 Fcylcold9
	565	66	78.6	840	10	CG237974 CG237974 OG1BM88TH
	566	66	78.6	893	8	CX050389 CX050389 UCRCS09_3
	567	66	78.6	897	6	CF828724 CF828724 UCRCR01_0
	568	66	78.6	903	10	DU067479 DU067479 77885 Tom
	569	66	78.6	904	8	CX049781 CX049781 UCRCS09_2
	570	66	78.6	911	10	DU052888 DU052888 140288 To
c	571	66	78.6	918	10	CZ986406 CZ986406 205005 To
c	572	66	78.6	929	7	CN645900 CN645900 ILLUMIGEN
c	573	66	78.6	1020	11	CNS07CMU AL439372 T7 end of
	574	66	78.6	1087	11	CNS03KSC AL248565 Tetraodon
c	575	66	78.6	1097	11	CNS050VO AL316077 Tetraodon
c	576	66	78.6	1101	11	CNS05SQY AL352195 Tetraodon
	577	66	78.6	1116	8	DR124377 DR124377 49161731
	578	65.5	78.0	545	5	BU002792 BU002792 QGG33A05.
c	579	65.5	78.0	691	10	CZ309630 CZ309630 ZMMBF0006
	580	65.5	78.0	929	9	CC364379 CC364379 PUHRB42TD
c	581	65.5	78.0	964	9	CC364375 CC364375 PUHRB42TB
	582	65	77.4	184	7	CV583709 CV583709 L_AH-aaa1
c	583	65	77.4	185	7	CV583842 CV583842 L_AH-aaa1
	584	65	77.4	211	4	AK185713 AK185713 Mus muscu
	585	65	77.4	219	3	BM164612 BM164612 EST567135
	586	65	77.4	240	1	AU074850 AU074850 AU074850
	587	65	77.4	262	10	CZ918340 CZ918340 4021009A0
	588	65	77.4	285	8	W40684 W40684 mc66a06.r1
c	589	65	77.4	308	5	BY356096 BY356096 BY356096
	590	65	77.4	315	1	AI483545 AI483545 EST249394
	591	65	77.4	321	7	CF985192 CF985192 24466rsic
	592	65	77.4	325	7	CF962054 CF962054 9000rsice
	593	65	77.4	334	2	BE192611 BE192611 EST331340
	594	65	77.4	341	5	BY114663 BY114663 BY114663
	595	65	77.4	344	5	BY104996 BY104996 BY104996
	596	65	77.4	347	1	AA098068 AA098068 mk18e01.r
	597	65	77.4	349	5	BY776640 BY776640 BY776640
	598	65	77.4	357	8	W62873 W62873 md88a07.r1
	599	65	77.4	358	9	CE252889 CE252889 tigr-gss-
	600	65	77.4	362	1	AA409731 AA409731 EST01462

ALIGNMENTS

```
RESULT 1
BJ340611/c
LOCUS       BJ340611                125 bp    mRNA    linear   EST 07-MAR-2002
DEFINITION  BJ340611 Dictyostelium discoideum cDNA library, AF Dictyostelium
            discoideum cDNA clone dda3c02 3', mRNA sequence.
ACCESSION   BJ340611
VERSION     BJ340611.1  GI:19248973
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 125)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the aggregation
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .125
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dda3c02"
                     /sex="mat A"
                     /dev_stage="Aggregation stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN

Alignment Scores:
Pred. No.:                0.0227          Length:          125
Score:                    84.00           Matches:         12
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       3               Gaps:            0

US-09-858-332G-15 (1-12) x BJ340611 (1-125)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        121 CACAATCACAATCACAATCACAATCACAATCACAAT 86

RESULT 2
BJ429957/c
LOCUS       BJ429957                174 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ429957 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv5d17 3', mRNA sequence.
ACCESSION   BJ429957
VERSION     BJ429957.1  GI:19404679
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
```

```
REFERENCE   1  (bases 1 to 174)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .174
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv5d17"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:                0.0321          Length:          174
Score:                    84.00           Matches:         12
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       3               Gaps:            0

US-09-858-332G-15 (1-12) x BJ429957 (1-174)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        127 CACAATCACAATCACAATCACAATCACAATCACAAT 92

RESULT 3
AU053723
LOCUS       AU053723                209 bp    mRNA    linear   EST 28-APR-1999
DEFINITION  AU053723 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
            discoideum cDNA clone SLJ555, mRNA sequence.
ACCESSION   AU053723
VERSION     AU053723.1  GI:4702205
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 209)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp
            PROJECT = Dictyostelium discoideum cDNA project in Japan.
FEATURES             Location/Qualifiers
     source          1. .209
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SLJ555"
                     /dev_stage="slug"
                     /clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
ORIGIN

Alignment Scores:
Pred. No.:                0.0389          Length:          209
Score:                    84.00           Matches:         12
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       1               Gaps:            0

US-09-858-332G-15 (1-12) x AU053723 (1-209)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db          2 CACAACCACAACCACAACCACAACCACAACCACAAC 37

RESULT 4
BQ586973
LOCUS       BQ586973                214 bp    mRNA    linear   EST 06-DEC-2002
DEFINITION  E012352-024-011-P08-SP6 MPIZ-ADIS-024-leaf Beta vulgaris cDNA clone
            024-011-P08 5-PRIME, mRNA sequence.
ACCESSION   BQ586973
VERSION     BQ586973.1  GI:26116555
KEYWORDS    EST.
SOURCE      Beta vulgaris
  ORGANISM  Beta vulgaris
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            Caryophyllales; Amaranthaceae; Beta.
REFERENCE   1  (bases 1 to 214)
  AUTHORS   Herwig,R., Schulz,B., Weisshaar,B., Hennig,S., Steinfath,M.,
            Drungowski,M., Stahl,D., Wruck,W., Menze,A., O'Brien,J., Lehrach,H.
            and Radelof,U.
  TITLE     Construction of a 'unigene' cDNA clone set by oligonucleotide
            fingerprinting allows access to 25 000 potential sugar beet genes
  JOURNAL   Plant J. 32 (5), 845-857 (2002)
  PUBMED    12472698
COMMENT     Contact: Weisshaar B
            ADIS DNA core facility at MPIZ
            Max-Planck-Institute for Plant Breeding Research
            Carl-von-Linne Weg 10, 50829 Koeln, Germany
            Fax: 00492215062851
            Email: weisshaa@mpiz-koeln.mpg.de
            Insert Length: 214   Std Error: 0.00
            Plate: 11  row: P  column: 08
            Seq primer: SP6; CATACGATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .214
                     /organism="Beta vulgaris"
                     /mol_type="mRNA"
                     /cultivar="KWS2320 (double haploid, monogerm breeding
                     line)"
                     /db_xref="GABI:185888"
                     /db_xref="taxon:161934"
                     /clone="024-011-P08"
                     /tissue_type="leaf"
                     /lab_host="EMDH10B"
                     /clone_lib="MPIZ-ADIS-024-leaf"
                     /note="Vector: pCMVSPORT6; Site_1: SalI; Site_2: NotI;
                     cDNA library from sugar beet, library provided by KWS
                     Kleinwanzlebener Saatzucht AG Einbeck, Germany, contact:
                     b.schulz@kws.de; cloning sites SalI-NotI, primer sites and
                     orientation:
                     SP6-SalI-CCACGCGTCCG-5prime-cDNA-polyA-CC-NotI-T7;  Note:
                     Sequencing granted in the context of the GABI-Beet
                     project, local PI: Dr. Katharina Schneider, coordinator:
                     Prof. Christian Jung;  Sequence submission managed by
                     RZPD/GABI-Primary database:http://gabi.rzpd.de"
ORIGIN

Alignment Scores:
Pred. No.:                0.0399          Length:          214
Score:                    84.00           Matches:         12
Percent Similarity:       100.00%         Conservative:    0
```

```
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     5             Gaps:          0

US-09-858-332G-15 (1-12) x BQ586973 (1-214)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        169 CACAACCACAACCACAACCACAACCACAACCACAAC 204

RESULT 5
AU053106
LOCUS       AU053106                264 bp    mRNA    linear   EST 28-APR-1999
DEFINITION  AU053106 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
            discoideum cDNA clone SLF703, mRNA sequence.
ACCESSION   AU053106
VERSION     AU053106.1  GI:4701589
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 264)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp
            PROJECT = Dictyostelium discoideum cDNA project in Japan.
FEATURES             Location/Qualifiers
     source          1. .264
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SLF703"
                     /dev_stage="slug"
                     /clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
ORIGIN

Alignment Scores:
Pred. No.:              0.0497        Length:        264
Score:                  84.00         Matches:       12
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     1             Gaps:          0

US-09-858-332G-15 (1-12) x AU053106 (1-264)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db         41 CACAATCACAATCACAATCACAATCACAATCACAAT 76

RESULT 6
BW876703/c
LOCUS       BW876703                297 bp    mRNA    linear   EST 24-MAY-2005
DEFINITION  BW876703 Amphioxus Branchiostoma floridae unpublished cDNA library,
            neurula whole animal Branchiostoma floridae cDNA clone bbne120h22
            5', mRNA sequence.
ACCESSION   BW876703
VERSION     BW876703.1  GI:66491380
KEYWORDS    EST.
SOURCE      Branchiostoma floridae (Florida lancelet)
  ORGANISM  Branchiostoma floridae
            Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
            Branchiostoma.
REFERENCE   1  (bases 1 to 297)
  AUTHORS   Yu,J., Holland,L.Z., Shin-i,T., Kohara,Y., Satou,Y. and Satoh,N.
  TITLE     Expressed genes in Branchiostoma floridae
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .297
                     /organism="Branchiostoma floridae"
                     /mol_type="mRNA"
                     /db_xref="taxon:7739"
                     /clone="bbne120h22"
                     /tissue_type="whole animal"
                     /dev_stage="neurula"
                     /clone_lib="Amphioxus Branchiostoma floridae unpublished
                     cDNA library, neurula whole animal"
ORIGIN

Alignment Scores:
Pred. No.:              0.0562        Length:        297
Score:                  84.00         Matches:       12
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     5             Gaps:          0

US-09-858-332G-15 (1-12) x BW876703 (1-297)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        175 CATAATCATAATCATAATCATAATCATAATCATAAT 140

RESULT 7
BF547803
LOCUS       BF547803                298 bp    mRNA    linear   EST 11-DEC-2000
DEFINITION  UI-R-E0-bw-f-10-0-UI.r1 UI-R-E0 Rattus norvegicus cDNA clone
            UI-R-E0-bw-f-10-0-UI 5', mRNA sequence.
ACCESSION   BF547803
VERSION     BF547803.1  GI:11638910
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 298)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
   PUBMED   8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road , 4156  MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            clones will be available through Research Genetics (www.resgen.com)
            This clone is also available through the I.M.A.G.E. Consortium at
            LLNL (info@image.llnl.gov). IMAGE ID= 1769930 The following
            repetitive elements were found in this cDNA sequence: 227-279,
            >(CAA)n#Simple_repeat
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .298
                     /organism="Rattus norvegicus"
```

/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-E0-bw-f-10-0-UI"
/dev_stage="embryonic"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-E0"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: NotI; Site_2: EcoRI; This library consists of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture."

ORIGIN

Alignment Scores:

Pred. No.:	0.0564	Length:	298
Score:	84.00	Matches:	12
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-15 (1-12) x BF547803 (1-298)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        239 CACAACCATAACCACAACCACAACCACAACCACAAC 274
```

RESULT 8
D47535

LOCUS	D47535 317 bp mRNA linear EST 03-APR-2002
DEFINITION	RICS13103A Rice green shoot Oryza sativa (japonica cultivar-group) cDNA, mRNA sequence.
ACCESSION	D47535
VERSION	D47535.1 GI:701244
KEYWORDS	EST.
SOURCE	Oryza sativa (japonica cultivar-group)
ORGANISM	Oryza sativa (japonica cultivar-group) Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE	1 (bases 1 to 317)
AUTHORS	Sasaki,T., Miyao,A. and Yamamoto,K.
TITLE	Rice cDNA from callus 1995
JOURNAL	Unpublished (1995)
COMMENT	Contact: Takuji Sasaki National Institute of Agrobiological Resources Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan Tel: 81-298-38-7441 Fax: 81-298-38-7468 Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/ PROJECT ='RGP'.

FEATURES Location/Qualifiers
source 1. .317
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone_lib="Rice green shoot"
/note="Green shoot (8 days old)"

ORIGIN

Alignment Scores:

Pred. No.:	0.0602	Length:	317
Score:	84.00	Matches:	12
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	8	Gaps:	0

US-09-858-332G-15 (1-12) x D47535 (1-317)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        151 CACAACCACAACCACAACCACAATCACAACCACAAT 186
```

RESULT 9
CE715092/c

LOCUS	CE715092 318 bp DNA linear GSS 29-SEP-2003
DEFINITION	tigr-gss-dog-17000369508467 Dog Library Canis familiaris genomic, genomic survey sequence.
ACCESSION	CE715092
VERSION	CE715092.1 GI:37034530
KEYWORDS	GSS.
SOURCE	Canis familiaris (dog)
ORGANISM	Canis familiaris Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE	1 (bases 1 to 318)
AUTHORS	Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K., Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and Venter,J.C.
TITLE	The dog genome: survey sequencing and comparative analysis
JOURNAL	Science 301 (5641), 1898-1903 (2003)
PUBMED	14512627
COMMENT	Contact: Kirkness EF The Institute for Genomic Research Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive, Rockville, MD 20850, USA Tel: 301-838-0200 Fax: 301-838-0208 Email: ekirknes@tigr.org Class: shotgun.

FEATURES Location/Qualifiers
source 1. .318
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site_1: BstXI; Libraries were prepared from peripheral blood"

ORIGIN

Alignment Scores:

Pred. No.:	0.0603	Length:	318
Score:	84.00	Matches:	12
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	10	Gaps:	0

US-09-858-332G-15 (1-12) x CE715092 (1-318)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        254 CATAATCATAATCATAATCATAATCATAATCATAAT 219
```

RESULT 10
AU061961

LOCUS	AU061961 319 bp mRNA linear EST 20-MAY-1999
DEFINITION	AU061961 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium discoideum cDNA clone SLG781, mRNA sequence.
ACCESSION	AU061961
VERSION	AU061961.1 GI:4883065
KEYWORDS	EST.
SOURCE	Dictyostelium discoideum
ORGANISM	Dictyostelium discoideum Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE	1 (bases 1 to 319)
AUTHORS	Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,

```
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp
            PROJECT = Dictyostelium discoideum cDNA project in Japan.
FEATURES             Location/Qualifiers
     source          1. .319
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SLG781"
                     /dev_stage="slug"
                     /clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
ORIGIN

Alignment Scores:
Pred. No.:              0.0605          Length:         319
Score:                  84.00           Matches:        12
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-15 (1-12) x AU061961 (1-319)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db         85 CACAACCACAACCACAACCACAACCACAACCACAAC 120

RESULT 11
BP803155/c
LOCUS       BP803155                375 bp    mRNA    linear   EST 15-FEB-2005
DEFINITION  BP803155 RAFL14 Arabidopsis thaliana cDNA clone RAFL23-27-N21 5',
            mRNA sequence.
ACCESSION   BP803155
VERSION     BP803155.1  GI:59839234
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 375)
  AUTHORS   Seki,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M., Akiyama,K.,
            Iida,K., Enju,A., Sakurai,T., Arakawa,T., Carninci,P., Fukuda,S.,
            Iida,J., Kawai,J., Sasaki,D., Shiraki,T., Hayashizaki,Y. and
            Shinozaki,K.
  TITLE     Large-Scale Analysis of RIKEN Arabidopsis Full-length cDNAs
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            An Arabidopsis full-length cDNA library was constructed essentially
            as reported previously (Seki et al., 1998, 2002).This clone is in a
            modified pBluescript vector.
            Please visit our web site (http://pfgweb.gsc.riken.jp and
            http://rarge.gsc.riken.jp) for further details.
FEATURES             Location/Qualifiers
     source          1. .375
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="RAFL23-27-N21"
                     /tissue_type="root"
                     /lab_host="DH10B"
                     /clone_lib="RAFL14"
                     /note="Site_1: BamHI; Site_2: SalI"
ORIGIN

Alignment Scores:
Pred. No.:              0.0717          Length:         375
Score:                  84.00           Matches:        12
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-858-332G-15 (1-12) x BP803155 (1-375)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        112 CACAATCACAATCACAATCACAATCACAATCACAAT 77

RESULT 12
BG669342
LOCUS       BG669342                413 bp    mRNA    linear   EST 30-APR-2001
DEFINITION  DRNAAED03 Rat DRG Library Rattus norvegicus cDNA clone DRNAAED03
            5', mRNA sequence.
ACCESSION   BG669342
VERSION     BG669342.1  GI:13891264
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   Xiao,H.S., Huang,Q.H., Zhang,F.X., Bao,L., Lu,Y.J., Guo,C.,
            Yang,L., Huang,W.J., Fu,G., Xu,S.H., Cheng,X.P., Yan,Q., Zhu,Z.D.,
            Zhang,X., Chen,Z., Han,Z.G. and Zhang,X.
  TITLE     Identification of gene expression profile of dorsal root ganglion
            in the rat peripheral axotomy model of neuropathic pain
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (12), 8360-8366 (2002)
   PUBMED   12060780
COMMENT     Contact: Zhang Xu
            Laboratory of Sensory System
            Institute of Neuroscience
            320 Yue Yang Road, Shanghai 200031, P.R.China
            Tel: 86-21-64748700-121
            Fax: 86-21-64713446
            Email: xu.zhang@ion.ac.cn
            This clone is also available at Chinese National Human Genome
            Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park,
            Pudong New Area, P.R.China. Please contact with Zhang Xu
            (xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)
            PCR PRimers
            FORWARD: T3
            BACKWARD: T7
            Seq primer: T3
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .413
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="DRNAAED03"
                     /sex="male"
                     /tissue_type="dorsal root ganglion"
                     /dev_stage="adult"
                     /clone_lib="Rat DRG Library"
ORIGIN

Alignment Scores:
```

```
Pred. No.:              0.0793          Length:         413
Score:                  84.00           Matches:        12
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-858-332G-15 (1-12) x BG669342 (1-413)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        213 CACAACCATAACCACAACCACAACCACAACCACAAC 248

RESULT 13
AU034405
LOCUS       AU034405                 417 bp    mRNA    linear   EST 28-APR-1999
DEFINITION  AU034405 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
            discoideum cDNA clone SLC829, mRNA sequence.
ACCESSION   AU034405
VERSION     AU034405.1  GI:3799829
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 417)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     The Dictyostelium developmental cDNA project: generation and
            analysis of expressed sequence tags from the first-finger stage of
            development
  JOURNAL   DNA Res. 5 (6), 335-340 (1998)
   PUBMED   10048482
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp
            PROJECT = 'Dictyostelium discoideum cDNA project in Japan'
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .417
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SLC829"
                     /dev_stage="slug"
                     /clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
ORIGIN

Alignment Scores:
Pred. No.:              0.0801          Length:         417
Score:                  84.00           Matches:        12
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-15 (1-12) x AU034405 (1-417)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        129 CACAATCACAATCACAATCACAATCACAATCACAAT 164

RESULT 14
AI008498/c
LOCUS       AI008498                 426 bp    mRNA    linear   EST 25-JAN-1999
DEFINITION  EST202949 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone
            REMAY09 3' end, mRNA sequence.
ACCESSION   AI008498
VERSION     AI008498.1  GI:3222330
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 426)
  AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
  TITLE     Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Lee, NH
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.
FEATURES             Location/Qualifiers
     source          1. .426
                     /organism="Rattus sp."
                     /mol_type="mRNA"
                     /db_xref="ATCC (inhost):2016683"
                     /db_xref="taxon:10118"
                     /clone="REMAY09"
                     /dev_stage="embryo 8, 12, 18 dpc"
                     /clone_lib="Normalized rat embryo, Bento Soares"
                     /note="Vector: pT7T3Pac; Site_1: EcoRI; Site_2: NotI"
ORIGIN

Alignment Scores:
Pred. No.:              0.0819          Length:         426
Score:                  84.00           Matches:        12
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-15 (1-12) x AI008498 (1-426)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        322 CACAACCATAACCACAACCACAACCACAACCACAAC 287

RESULT 15
AU033706
LOCUS       AU033706                 427 bp    mRNA    linear   EST 28-APR-1999
DEFINITION  AU033706 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
            discoideum cDNA clone SLB340, mRNA sequence.
ACCESSION   AU033706
VERSION     AU033706.1  GI:3799130
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 427)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     The Dictyostelium developmental cDNA project: generation and
            analysis of expressed sequence tags from the first-finger stage of
            development
  JOURNAL   DNA Res. 5 (6), 335-340 (1998)
   PUBMED   10048482
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp
```

```
           PROJECT = 'Dictyostelium discoideum cDNA project in Japan'
           POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .427
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SLB340"
                     /dev_stage="slug"
                     /clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
ORIGIN

Alignment Scores:
Pred. No.:               0.0821          Length:         427
Score:                   84.00           Matches:        12
Percent Similarity:      100.00%         Conservative:   0
Best Local Similarity:   100.00%         Mismatches:     0
Query Match:             100.00%         Indels:         0
DB:                      1               Gaps:           0

US-09-858-332G-15 (1-12) x AU033706 (1-427)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        207 CACAATCACAATCACAATCACAATCACAATCACAAT 242

RESULT 16
AU033893
LOCUS       AU033893                427 bp    mRNA    linear   EST 28-APR-1999
DEFINITION  AU033893 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
            discoideum cDNA clone SLB578, mRNA sequence.
ACCESSION   AU033893
VERSION     AU033893.1  GI:3799317
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 427)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     The Dictyostelium developmental cDNA project: generation and
            analysis of expressed sequence tags from the first-finger stage of
            development
  JOURNAL   DNA Res. 5 (6), 335-340 (1998)
   PUBMED   10048482
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp
            PROJECT = 'Dictyostelium discoideum cDNA project in Japan'
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .427
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SLB578"
                     /dev_stage="slug"
                     /clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
ORIGIN

Alignment Scores:
Pred. No.:               0.0821          Length:         427
Score:                   84.00           Matches:        12
Percent Similarity:      100.00%         Conservative:   0
Best Local Similarity:   100.00%         Mismatches:     0
Query Match:             100.00%         Indels:         0
DB:                      1               Gaps:           0

US-09-858-332G-15 (1-12) x AU033893 (1-427)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        207 CACAATCACAATCACAATCACAATCACAATCACAAT 242

RESULT 17
BJ434889/c
LOCUS       BJ434889                446 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ434889 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv25f07 3', mRNA sequence.
ACCESSION   BJ434889
VERSION     BJ434889.1  GI:19409611
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 446)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .446
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv25f07"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:               0.0859          Length:         446
Score:                   84.00           Matches:        12
Percent Similarity:      100.00%         Conservative:   0
Best Local Similarity:   100.00%         Mismatches:     0
Query Match:             100.00%         Indels:         0
DB:                      3               Gaps:           0

US-09-858-332G-15 (1-12) x BJ434889 (1-446)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        149 CACAATCACAATCACAATCACAATCACAATCACAAT 114

RESULT 18
AU040068
LOCUS       AU040068                453 bp    mRNA    linear   EST 29-MAR-1999
DEFINITION  AU040068 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
            discoideum cDNA clone SLA355, mRNA sequence.
ACCESSION   AU040068
VERSION     AU040068.1  GI:4009308
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 453)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     The Dictyostelium developmental cDNA project: generation and
```

```
            analysis of expressed sequence tags from the first-finger stage of
            development
  JOURNAL   DNA Res. 5 (6), 335-340 (1998)
    PUBMED  10048482
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp
            PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1. .453
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SLA355"
                     /dev_stage="slug"
                     /clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
ORIGIN

Alignment Scores:
Pred. No.:                0.0874          Length:          453
Score:                    84.00           Matches:         12
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       1               Gaps:            0

US-09-858-332G-15 (1-12) x AU040068 (1-453)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        233 CACAATCACAATCACAATCACAATCACAATCACAAT 268

RESULT 19
AU036977
LOCUS       AU036977                459 bp    mRNA    linear   EST 29-MAR-1999
DEFINITION  AU036977 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
            discoideum cDNA clone SSB896, mRNA sequence.
ACCESSION   AU036977
VERSION     AU036977.1  GI:3983730
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 459)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     The Dictyostelium developmental cDNA project: generation and
            analysis of expressed sequence tags from the first-finger stage of
            development
  JOURNAL   DNA Res. 5 (6), 335-340 (1998)
    PUBMED  10048482
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp
            PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1. .459
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SSB896"
                     /dev_stage="slug"
                     /clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
ORIGIN

Alignment Scores:
Pred. No.:                0.0886          Length:          459
Score:                    84.00           Matches:         12
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       1               Gaps:            0

US-09-858-332G-15 (1-12) x AU036977 (1-459)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        237 CACAATCACAATCACAATCACAATCACAATCACAAT 272

RESULT 20
CB714934
LOCUS       CB714934                466 bp    mRNA    linear   EST 10-APR-2003
DEFINITION  AMGNNUC:MRPE3-00062-C4-A placenta embryo D17 (10379) Rattus
            norvegicus cDNA clone mrpe3-00062-c4 5', mRNA sequence.
ACCESSION   CB714934
VERSION     CB714934.1  GI:29772082
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 466)
  AUTHORS   Amgen EST Program.
  TITLE     Amgen Rat EST Program
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Dan Fitzpatrick
            Amgen, Inc
            One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA
            Tel: 805 447-4881
            Plate: 00062  row: c  column: 4.
FEATURES             Location/Qualifiers
     source          1. .466
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
                     /clone="mrpe3-00062-c4"
                     /tissue_type="placenta embryo"
                     /clone_lib="placenta embryo D17 (10379)"
                     /note="Vector: pSPORT1; placenta embryo D17"
ORIGIN

Alignment Scores:
Pred. No.:                0.09            Length:          466
Score:                    84.00           Matches:         12
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       6               Gaps:            0

US-09-858-332G-15 (1-12) x CB714934 (1-466)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        405 CACAACCACAACCACAACCACAACCACAACCACAAC 440

RESULT 21
BJ431306/c
LOCUS       BJ431306                467 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ431306 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv13o13 3', mRNA sequence.
ACCESSION   BJ431306
VERSION     BJ431306.1  GI:19406028
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
```

```
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 467)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .467
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv13o13"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:              0.0902          Length:         467
Score:                  84.00           Matches:        12
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-858-332G-15 (1-12) x BJ431306 (1-467)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        211 CACAATCACAATCACAATCACAATCACAATCACAAT 176

RESULT 22
BJ429835/c
LOCUS       BJ429835                472 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ429835 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv5d06 3', mRNA sequence.
ACCESSION   BJ429835
VERSION     BJ429835.1  GI:19404557
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 472)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .472
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv5d06"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:              0.0912          Length:         472
Score:                  84.00           Matches:        12
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-858-332G-15 (1-12) x BJ429835 (1-472)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        161 CACAATCACAATCACAATCACAATCACAATCACAAT 126

RESULT 23
BJ431249/c
LOCUS       BJ431249                476 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ431249 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv13o07 3', mRNA sequence.
ACCESSION   BJ431249
VERSION     BJ431249.1  GI:19405971
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 476)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .476
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv13o07"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:              0.092           Length:         476
Score:                  84.00           Matches:        12
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-858-332G-15 (1-12) x BJ431249 (1-476)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        210 CACAATCACAATCACAATCACAATCACAATCACAAT 175

RESULT 24
CR294373/c
LOCUS       CR294373                480 bp    DNA     linear   GSS 19-NOV-2004
DEFINITION  mte1-10O13FM1 BAC end, cultivar Jemalong A17 of Medicago
            truncatula, genomic survey sequence.
ACCESSION   CR294373
VERSION     CR294373.1  GI:44684797
KEYWORDS    GSS.
```

```
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 480)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
FEATURES             Location/Qualifiers
     source          1. .480
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /cultivar="Jemalong A17"
                     /db_xref="taxon:3880"
                     /clone_lib="MTE1"
                     /note="Vector: pIndigoBAC ; Site_1: EcoRI ; Site_2: EcoRI
                     ; Debelle F. and Chalhoub B.
                     Genoscope sequence ID : mte1-10O13FM1"
ORIGIN

Alignment Scores:
Pred. No.:                0.0928          Length:          480
Score:                    84.00           Matches:         12
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       11              Gaps:            0

US-09-858-332G-15 (1-12) x CR294373 (1-480)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        274 CACAATCATAATCATAATCATAATCACAATCACAAT 239

RESULT 25
AZ321550
LOCUS       AZ321550                 488 bp    DNA     linear   GSS 29-SEP-2000
DEFINITION  1M0042I11F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0042I11 F, genomic survey sequence.
ACCESSION   AZ321550
VERSION     AZ321550.1  GI:10374406
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 488)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
            Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
            Niederhausern,A. and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0042  row: I  column: 11
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 488.
FEATURES             Location/Qualifiers
     source          1. .488
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0042I11"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
ORIGIN

Alignment Scores:
Pred. No.:                0.0944          Length:          488
Score:                    84.00           Matches:         12
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       9               Gaps:            0

US-09-858-332G-15 (1-12) x AZ321550 (1-488)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db         55 CACAACCACAACCACAACCACAACCACAACCACAAC 90

RESULT 26
BJ388347
LOCUS       BJ388347                 494 bp    mRNA    linear   EST 08-MAR-2002
DEFINITION  BJ388347 Dictyostelium discoideum cDNA library, SF Dictyostelium
            discoideum cDNA clone dds9h04 5', mRNA sequence.
ACCESSION   BJ388347
VERSION     BJ388347.1  GI:19299433
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 494)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the slug stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .494
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dds9h04"
```

```
                     /sex="mat A"
                     /dev_stage="Slug stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, SF"
ORIGIN

Alignment Scores:
Pred. No.:              0.0956          Length:         494
Score:                  84.00           Matches:        12
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-858-332G-15 (1-12) x BJ388347 (1-494)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        400 CACAACCACAACCACAACCACAACCACAACCACAAC 435

RESULT 27
C90140
LOCUS       C90140                    496 bp    mRNA    linear   EST 20-APR-1998
DEFINITION  C90140 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
            discoideum cDNA clone SSI136, mRNA sequence.
ACCESSION   C90140
VERSION     C90140.1  GI:3059760
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 496)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .496
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SSI136"
                     /dev_stage="slug"
                     /clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
ORIGIN

Alignment Scores:
Pred. No.:              0.096           Length:         496
Score:                  84.00           Matches:        12
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     5               Gaps:           0

US-09-858-332G-15 (1-12) x C90140 (1-496)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        201 CACAATCACAATCACAATCACAATCACAATCACAAT 236

RESULT 28
CK041708
LOCUS       CK041708                  500 bp    mRNA    linear   EST 05-FEB-2005
DEFINITION  48024rsiceh_17347.y1 Oryza sativa cv. PA64s panicle ferile cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CK041708
VERSION     CK041708.1  GI:58653028
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 500
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .500
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="PA64s"
                     /db_xref="taxon:39946"
                     /tissue_type="panicle"
                     /cell_type="ferile"
                     /dev_stage="heading/flowering"
                     /clone_lib="Oryza sativa cv. PA64s panicle ferile cDNA
                     library"
ORIGIN

Alignment Scores:
Pred. No.:              0.0969          Length:         500
Score:                  84.00           Matches:        12
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     7               Gaps:           0

US-09-858-332G-15 (1-12) x CK041708 (1-500)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        217 CACAACCACAACCACAACCACAATCACAACCACAAT 252

RESULT 29
CK618125
LOCUS       CK618125                  504 bp    mRNA    linear   EST 26-JAN-2004
DEFINITION  ml05h06.y1 Mouse retina, unamplified: mk/ml Mus musculus cDNA clone
            ml05h06 5', mRNA sequence.
ACCESSION   CK618125
VERSION     CK618125.1  GI:41339011
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 504)
  AUTHORS   Ida,H., Boylan,S., Weigel,A., Smit-McBride,Z., Chao,A., Gao,J.,
            Buchoff,P., Wistow,G. and Hjelmeland,L.
```

```
  TITLE     Expressed sequence tag analysis of mouse retina
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Wistow G
            Section on Molecular Structure and Function
            National Eye Institute
            6/331, NIH, Bethesda, MD 20892-2740, USA
            Tel: 301 402 3452
            Fax: 301 496 0078
            Email: graeme@helix.nih.gov
            Plate: 05  row: h  column: 06
            Seq primer: M13RP1 reverse primer (ABI).
FEATURES             Location/Qualifiers
     source          1. .504
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57Bl6J"
                     /db_xref="taxon:10090"
                     /clone="ml05h06"
                     /sex="Male"
                     /tissue_type="retina"
                     /dev_stage="Adult"
                     /lab_host="EMDH10B"
                     /clone_lib="Mouse retina, unamplified: mk/ml"
                     /note="Organ: Eye; Vector: pSport1; 270ug total RNA was
                     extracted from 200 adult male mouse retinas. A
                     directionally cloned cDNA library in the pSPORT1
                     vector(Life Technologies) was constructed at Bioserve
                     Biotechnology (Laurel MD) essentially following the
                     protocols of the SuperScript Plasmid System full details
                     of which are contained in the manufacturer's Instruction
                     manual (http://www.lifetech.com/). First strand synthesis
                     was carried out using a Not I primer-adapter
                     [5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC(T)15-3']. cDNA was
                     cloned in Not I/Sal I sites. EST analysis was performed on
                     the unamplified library at the NIH Intramural Sequencing
                     Center (NISC)."
ORIGIN

Alignment Scores:
Pred. No.:                0.0977          Length:          504
Score:                    84.00           Matches:         12
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       7               Gaps:            0

US-09-858-332G-15 (1-12) x CK618125 (1-504)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        109 CACAACCACAACCACAACCACAACCACAACCACAAC 144

RESULT 30
CB264620/c
LOCUS       CB264620                528 bp    mRNA    linear   EST 06-NOV-2003
DEFINITION  50-E014659-035-002-C14-T7R MPIZ-ADIS-035 Arabidopsis thaliana cDNA
            clone MPIZp2000C142Q 5-PRIME, mRNA sequence.
ACCESSION   CB264620
VERSION     CB264620.1  GI:32889393
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 528)
  AUTHORS   Schmid,K.J., Soerensen,T.R., Stracke,R., Torjek,O., Altmann,T.,
            Mitchell-Olds,T. and Weisshaar,B.
  TITLE     Large-scale identification and analysis of genome-wide
            single-nucleotide polymorphisms for mapping in Arabidopsis thaliana
  JOURNAL   Genome Res. 13 (6), 1250-1257 (2003)
   PUBMED   12799357
COMMENT     Contact: Weisshaar B
            ADIS DNA core facility at MPIZ
            Max-Planck-Institute for Plant Breeding Research
            Carl-von-Linne Weg 10, 50829 Koeln, Germany
            Fax: 00492215062851
            Email: weisshaa@mpiz-koeln.mpg.de
            Insert Length: 528   Std Error: 0.00
            Plate: 2  row: C  column: 14
            Seq primer: T7R; CTAATACGACTCACTATAGGGA.
FEATURES             Location/Qualifiers
     source          1. .528
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /cultivar="Achkarren-2 (Ak-2)"
                     /db_xref="GABI:594788"
                     /db_xref="taxon:3702"
                     /clone="MPIZp2000C142Q"
                     /tissue_type="inflorescence"
                     /lab_host="E. coli TOP10"
                     /clone_lib="MPIZ-ADIS-035"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: Not1; cDNA
                     library from Arabidopsis thaliana, accession Achkarren-2;
                     inflorescences from flower buds to young siliques; library
                     was made at the Max-Planck-Institute for Plant Breeding
                     Research, Cologne, Germany; cloning sites SalI-NotI,
                     primer sites and orientation:
                     T7-SalI-CCACGCGTCCG-5prime-cDNA-polyA-CC-NotI-SP6; Note:
                     Sequencing granted in the context of the GABI Arabidopsis
                     VerbundI: Genetic Diversity, 'Establishment of
                     high-efficiency SNP-based mapping tools and development of
                     methods for genome-wide mutation detection' PI: Bernd
                     Weisshaar; Sequence submission managed by
                     RZPD/GABI-Primary database: http://gabi.rzpd.de This clone
                     is available from RZPD; contact RZPD (clone@rzpd.de) for
                     further information."
ORIGIN

Alignment Scores:
Pred. No.:                0.103           Length:          528
Score:                    84.00           Matches:         12
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       6               Gaps:            0

US-09-858-332G-15 (1-12) x CB264620 (1-528)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db         98 CACAATCACAATCACAATCACAATCACAATCACAAT 63

RESULT 31
AQ491972
LOCUS       AQ491972                533 bp    DNA     linear   GSS 24-APR-1999
DEFINITION  V100B10 mTn-3xHA/lacZ Insertion Library Saccharomyces cerevisiae
            genomic 5', genomic survey sequence.
ACCESSION   AQ491972
VERSION     AQ491972.1  GI:4677846
KEYWORDS    GSS.
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 533)
  AUTHORS   Ross-Macdonald,P., Roemer,T., Coelho,P.S.R., Agarwal,S., Kumar,A.,
            desEtages,S.A., Cheung,K.-H., Sheehan,A., Symoniatis,D., Jansen,R.,
            Umansky,L., Heidtman,M., Nelson,K., Iwasaki,H., Kanada,D., Lugo,R.,
            Hager,K., Miller,P., Roeder,G.S. and Snyder,M.
  TITLE     Large-Scale Analysis of the Yeast Genome by Transposon Tagging and
            Gene Disruption
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Kumar A
            Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
            Yale University
```

```
          P.O. Box 208103, New Haven, CT 06520-8103, USA
          Tel: 203 432 9949
          Fax: 203 432 6161
          Email: anuj.kumar@yale.edu
          Sequence corresponds to the extreme 5' end of mTn-3xHA/lacZ (ending
          in TCAGACCCC) followed by native Saccharomyces cerevisiae genomic
          sequence immediately upstream; generated by single read using
          sequencing primer indicated above; enables determination of site of
          mTn-3xHA/lacZ insertion.
          Seq primer: GGCCTTCTTTCTTTGGAAGTAC
          Class: transposon-tagged.
FEATURES             Location/Qualifiers
     source          1. .533
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4932"
                     /lab_host="E. coli"
                     /clone_lib="mTn-3xHA/lacZ Insertion Library"
                     /note="Vector: pHSS6-Sal; A yeast genomic DNA library
                     (lacking mitochondrial DNA) was prepared in pHSS6-Sal;
                     genomic DNA was size-fractionated (DNA of roughly 2-3 kb
                     in length) prior to cloning.  This library was
                     subsequently mutagenized with a mTn-3xHA/lacZ
                     minitransposon containing lacZ, URA3, and tet resistance."
ORIGIN

Alignment Scores:
Pred. No.:                0.104          Length:          533
Score:                    84.00          Matches:         12
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       9              Gaps:            0

US-09-858-332G-15 (1-12) x AQ491972 (1-533)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        109 CATAATCATAATCATAATCATAATCATAATCATAAT 144

RESULT 32
C22962
LOCUS       C22962                  538 bp    mRNA    linear   EST 02-APR-1997
DEFINITION  C22962 Dictyostelium discoideum FC (H.Urushihara) Dictyostelium
            discoideum cDNA clone FC-AL02, mRNA sequence.
ACCESSION   C22962
VERSION     C22962.1  GI:1922111
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 538)
  AUTHORS   Shimizu,H. and Urushihara,H.
  TITLE     Sexual-cDNA in CSM
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .538
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="KAX3"
                     /db_xref="taxon:44689"
                     /clone="FC-AL02"
                     /sex="mat A"
                     /tissue_type="gamete"
                     /dev_stage="Sexual"
                     /clone_lib="Dictyostelium discoideum FC (H.Urushihara)"
ORIGIN

Alignment Scores:
Pred. No.:                0.105          Length:          538
Score:                    84.00          Matches:         12
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       5              Gaps:            0

US-09-858-332G-15 (1-12) x C22962 (1-538)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        460 CACAACCACAACCACAACCACAACCACAACCACAAC 495

RESULT 33
AA874794/c
LOCUS       AA874794                540 bp    mRNA    linear   EST 04-JUL-1999
DEFINITION  UI-R-E0-bw-f-10-0-UI.s1 UI-R-E0 Rattus norvegicus cDNA clone
            UI-R-E0-bw-f-10-0-UI 3' similar to gi|189378|gb|M38188|HUMOGC Human
            unknown protein from clone pHGR74 mRNA, complete cds, mRNA
            sequence.
ACCESSION   AA874794
VERSION     AA874794.1  GI:4230677
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 540)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
   PUBMED   8889548
COMMENT     On Mar 20, 1998 this sequence version replaced gi:2979742.
            Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road , 4156  MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            The sequence tag present in the cDNA between the NotI site and the
            oligo-dT track served to identify it as a clone from the normalized
            adult 8-Day-Embryo library. cDNA Library Preparation: M. Fatima
            Bonaldo, Ph.D. Clone distribution: clones will be available through
            Research Genetics This clone is also available through the
            I.M.A.G.E. Consortium at LLNL (info@image.llnl.gov). IMAGE
            ID=1769930
            Seq primer: M13 Forward
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .540
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-E0-bw-f-10-0-UI"
                     /dev_stage="embryonic"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="UI-R-E0"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: NotI; Site_2: EcoRI; This library
                     consists of a mixture of individually tagged normalized
                     libraries constructed from 8, 12 and 18-day embryo.  The
                     tag is a string  of 3-5 nucleotides present between the
                     Not I site and the oligo-dT track  which allows
                     identification of the library of origin of a clone within
                     the mixture."
ORIGIN
```

```
Alignment Scores:
Pred. No.:                0.105          Length:          540
Score:                    84.00          Matches:         12
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       1              Gaps:            0

US-09-858-332G-15 (1-12) x AA874794 (1-540)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        537 CATAACCACAACCACAACCACAACCACAACCACAAT 502

RESULT 34
BJ429095/c
LOCUS       BJ429095                546 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ429095 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv2i07 3', mRNA sequence.
ACCESSION   BJ429095
VERSION     BJ429095.1  GI:19403817
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 546)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .546
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv2i07"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:                0.106          Length:          546
Score:                    84.00          Matches:         12
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       3              Gaps:            0

US-09-858-332G-15 (1-12) x BJ429095 (1-546)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        166 CACAATCACAATCACAATCACAATCACAATCACAAT 131

RESULT 35
BJ430064/c
LOCUS       BJ430064                556 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ430064 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv5o23 3', mRNA sequence.
ACCESSION   BJ430064
VERSION     BJ430064.1  GI:19404786
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 556)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .556
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv5o23"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:                0.108          Length:          556
Score:                    84.00          Matches:         12
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       3              Gaps:            0

US-09-858-332G-15 (1-12) x BJ430064 (1-556)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        180 CACAATCACAATCACAATCACAATCACAATCACAAT 145

RESULT 36
BG375827/c
LOCUS       BG375827                559 bp    mRNA    linear   EST 12-MAR-2001
DEFINITION  UI-R-CS0-btf-h-04-0-UI.s1 UI-R-CS0 Rattus norvegicus cDNA clone
            UI-R-CS0-btf-h-04-0-UI 3', mRNA sequence.
ACCESSION   BG375827
VERSION     BG375827.1  GI:13300299
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 559)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  PUBMED    8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road , 4156  MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            normalized rat heart pool library cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: clones will be available through
```

```
          Research Genetics (www.resgen.com) The following repetitive
          elements were found in this cDNA sequence: 504-556,
          >(CAA)n#Simple_repeat
          Seq primer: M13 Forward
          POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .559
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-CS0-btf-h-04-0-UI"
                     /dev_stage="ADULT"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="UI-R-CS0"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CS0
                     library is a normalized library constructed from the
                     following rat heart tissues: embryonic day 17, embryonic
                     day 19, embryonic day 21, adult day 1, adult day 12, adult
                     day 75, adult day 200. For a detailed description of the
                     library from which this clone was derived, please visit
                     our web site at ratest.eng.uiowa.edu. The subtraction has
                     been previously described in (Bonaldo, Lennon and Soares,
                     Genome Research 6:791-806, 1996)
                     TAG_TISSUE=rat heart pool
                     TAG_LIB=UI-R-CS0
                     TAG_SEQ=ATAAGATAAC"
ORIGIN

Alignment Scores:
Pred. No.:                0.109          Length:         559
Score:                    84.00          Matches:        12
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       2              Gaps:           0

US-09-858-332G-15 (1-12) x BG375827 (1-559)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        544 CACAACCATAACCACAACCACAACCACAACCACAAC 509

RESULT 37
AU265802
LOCUS       AU265802                560 bp    mRNA    linear   EST 26-APR-2004
DEFINITION  AU265802 VS Dictyostelium discoideum cDNA clone VSF793 5', mRNA
            sequence.
ACCESSION   AU265802
VERSION     AU265802.1  GI:20524600
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 560)
  AUTHORS   Urushihara,H., Morio,T., Saito,T., Kohara,Y., Koriki,E., Ochiai,H.,
            Maeda,M., Williams,J.G., Takeuchi,I. and Tanaka,Y.
  TITLE     Analyses of cDNAs from growth and slug stages of Dictyostelium
            discoideum
  JOURNAL   Nucleic Acids Res. 32 (5), 1647-1653 (2004)
   PUBMED   15010511
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .560
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="VSF793"
                     /sex="mat A"
                     /dev_stage="vegetative"
                     /clone_lib="VS"
ORIGIN

Alignment Scores:
Pred. No.:                0.109          Length:         560
Score:                    84.00          Matches:        12
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       1              Gaps:           0

US-09-858-332G-15 (1-12) x AU265802 (1-560)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        413 CACAACCACAACCACAACCACAACCACAACCACAAC 448

RESULT 38
BJ329834
LOCUS       BJ329834                560 bp    mRNA    linear   EST 05-MAR-2002
DEFINITION  BJ329834 Dictyostelium discoideum cDNA library, AF Dictyostelium
            discoideum cDNA clone dda26o07 5', mRNA sequence.
ACCESSION   BJ329834
VERSION     BJ329834.1  GI:19159964
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 560)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the aggregation
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .560
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dda26o07"
                     /sex="mat A"
                     /dev_stage="Aggregation stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN

Alignment Scores:
Pred. No.:                0.109          Length:         560
Score:                    84.00          Matches:        12
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       3              Gaps:           0

US-09-858-332G-15 (1-12) x BJ329834 (1-560)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        202 CACAATCACAATCACAATCACAATCACAATCACAAT 237

RESULT 39
BJ430176/c
```

```
LOCUS       BJ430176                571 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ430176 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv6h09 3', mRNA sequence.
ACCESSION   BJ430176
VERSION     BJ430176.1  GI:19404898
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 571)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .571
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv6h09"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:                0.111          Length:         571
Score:                    84.00          Matches:        12
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       3              Gaps:           0

US-09-858-332G-15 (1-12) x BJ430176 (1-571)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        171 CACAATCACAATCACAATCACAATCACAATCACAAT 136

RESULT 40
BJ432285/c
LOCUS       BJ432285                571 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ432285 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv18b10 3', mRNA sequence.
ACCESSION   BJ432285
VERSION     BJ432285.1  GI:19407007
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 571)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .571
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv18b10"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:                0.111          Length:         571
Score:                    84.00          Matches:        12
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       3              Gaps:           0

US-09-858-332G-15 (1-12) x BJ432285 (1-571)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        194 CACAATCACAATCACAATCACAATCACAATCACAAT 159

RESULT 41
BU886048/c
LOCUS       BU886048                577 bp    mRNA    linear   EST 17-OCT-2002
DEFINITION  R039F11 Populus root cDNA library Populus tremula x Populus
            tremuloides cDNA 5 prime, mRNA sequence.
ACCESSION   BU886048
VERSION     BU886048.1  GI:24077565
KEYWORDS    EST.
SOURCE      Populus tremula x Populus tremuloides
  ORGANISM  Populus tremula x Populus tremuloides
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Malpighiales; Salicaceae; Saliceae; Populus.
REFERENCE   1  (bases 1 to 577)
  AUTHORS   Unneberg,P., Bhalerao,R.R., Jansson,S. and Sterky,F.
  TITLE     The poplar tree transcriptome: Analysis of expressed sequence tags
            from multiple libraries
  JOURNAL   Unpublished (2002)
COMMENT     Contact: BHALERAO RUPALI R.
            Umea Plant Science Center
            Department of Plant Physiology
            University of Umea, 901 87 Umea, Sweden
            Tel: +46 90 786 5279
            Fax: +46 90 786 6676
            Email: rupali.bhalerao@plantphys.umu.se.
FEATURES             Location/Qualifiers
     source          1. .577
                     /organism="Populus tremula x Populus tremuloides"
                     /mol_type="mRNA"
                     /db_xref="taxon:47664"
                     /tissue_type="root"
                     /clone_lib="Populus root cDNA library"
ORIGIN

Alignment Scores:
Pred. No.:                0.112          Length:         577
Score:                    84.00          Matches:        12
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       5              Gaps:           0

US-09-858-332G-15 (1-12) x BU886048 (1-577)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        272 CATAATCATAATCATAATCATAATCATAATCATAAT 237

RESULT 42
CB255238/c
```

```
LOCUS       CB255238                579 bp    mRNA    linear   EST 31-DEC-2004
DEFINITION  46-E015331-019-005-K12-T7R MPIZ-ADIS-019 Arabidopsis thaliana cDNA
            clone MPIZp768K125Q 5-PRIME, mRNA sequence.
ACCESSION   CB255238
VERSION     CB255238.1  GI:56915863
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 579)
  AUTHORS   Jakoby,M., Stracke,R., Soerensen,T.R. and Weisshaar,B.
  TITLE     Arabidopsis thaliana cDNA library enriched in transcription factors
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Weisshaar B
            ADIS DNA core facility at MPIZ
            Max-Planck-Institute for Plant Breeding Research
            Carl-von-Linne Weg 10, 50829 Koeln, Germany
            Fax: 00492215062851
            Email: weisshaa@mpiz-koeln.mpg.de
            Insert Length: 579   Std Error: 0.00
            Plate: 5  row: K  column: 12
            Seq primer: T7R; CTAATACGACTCACTATAGGGA.
FEATURES             Location/Qualifiers
     source          1. .579
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /cultivar="At7"
                     /db_xref="GABI:597149"
                     /db_xref="taxon:3702"
                     /clone="MPIZp768K125Q"
                     /tissue_type="hypocotyl"
                     /dev_stage="tissue culture"
                     /lab_host="E. coli DH5alpha mcr"
                     /clone_lib="MPIZ-ADIS-019"
                     /note="Vector: pSPORT1; Site_1: NotI primer adapter;
                     Site_2: SalI primer adapter; RNA from cellculture (At7) 5
                     days after inoculation treated with 0.002 mM cycloheximid
                     for 2 h in the dark. Sequence submission managed by
                     RZPD/GABI-Primary database: http://gabi.rzpd.de. This
                     clone is available from RZPD; contact RZPD (clone@rzpd.de)
                     for further information."
ORIGIN

Alignment Scores:
Pred. No.:              0.113          Length:         579
Score:                  84.00          Matches:        12
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     6              Gaps:           0

US-09-858-332G-15 (1-12) x CB255238 (1-579)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        111 CACAATCACAATCACAATCACAATCACAATCACAAT 76

RESULT 43
BJ369186
LOCUS       BJ369186                582 bp    mRNA    linear   EST 08-MAR-2002
DEFINITION  BJ369186 Dictyostelium discoideum cDNA library, CF Dictyostelium
            discoideum cDNA clone ddc49k17 5', mRNA sequence.
ACCESSION   BJ369186
VERSION     BJ369186.1  GI:19278569
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 582)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the culmination
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .582
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddc49k17"
                     /sex="mat A"
                     /dev_stage="Culmination stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, CF"
ORIGIN

Alignment Scores:
Pred. No.:              0.114          Length:         582
Score:                  84.00          Matches:        12
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     3              Gaps:           0

US-09-858-332G-15 (1-12) x BJ369186 (1-582)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        284 CACAACCACAACCACAACCACAACCACAACCACAAC 319

RESULT 44
CB586337
LOCUS       CB586337                583 bp    mRNA    linear   EST 03-APR-2003
DEFINITION  AMGNNUC:NRDG1-00072-H8-A nrdg1 (10855) Rattus norvegicus cDNA clone
            nrdg1-00072-h8 5', mRNA sequence.
ACCESSION   CB586337
VERSION     CB586337.1  GI:29530378
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 583)
  AUTHORS   Amgen EST Program.
  TITLE     Amgen Rat EST Program
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Dan Fitzpatrick
            Amgen, Inc
            One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA
            Tel: 805 447-4881
            Plate: 00072  row: h  column: 8.
FEATURES             Location/Qualifiers
     source          1. .583
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
                     /clone="nrdg1-00072-h8"
                     /tissue_type="Dorsal Root Ganglia"
                     /clone_lib="nrdg1 (10855)"
                     /note="Vector: pSPORT1; Site_1: Sal1; Site_2: Not1; rat
                     dorsal root ganglia"
ORIGIN

Alignment Scores:
Pred. No.:              0.114          Length:         583
Score:                  84.00          Matches:        12
Percent Similarity:     100.00%        Conservative:   0
```

```
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     6              Gaps:          0

US-09-858-332G-15 (1-12) x CB586337 (1-583)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        296 CACAACCATAACCACAACCACAACCACAACCACAAC 331

RESULT 45
BP796399/c
LOCUS       BP796399                592 bp    mRNA    linear   EST 14-FEB-2005
DEFINITION  BP796399 RAFL5 Arabidopsis thaliana cDNA clone RAFL05-04-N22 5',
            mRNA sequence.
ACCESSION   BP796399
VERSION     BP796399.1  GI:59791943
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 592)
  AUTHORS   Seki,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M., Akiyama,K.,
            Iida,K., Enju,A., Sakurai,T., Arakawa,T., Carninci,P., Fukuda,S.,
            Iida,J., Kawai,J., Sasaki,D., Shiraki,T., Hayashizaki,Y. and
            Shinozaki,K.
  TITLE     Large-Scale Analysis of RIKEN Arabidopsis Full-length cDNAs
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            An Arabidopsis full-length cDNA library was constructed essentially
            as reported previously (Seki et al., 1998, 2002).This clone is in a
            modified pBluescript vector.
            Please visit our web site (http://pfgweb.gsc.riken.jp and
            http://rarge.gsc.riken.jp) for further details.
FEATURES             Location/Qualifiers
     source          1. .592
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="RAFL05-04-N22"
                     /dev_stage="rosette plants"
                     /lab_host="SOLR"
                     /clone_lib="RAFL5"
                     /note="Site_1: SstI; Site_2: XhoI; subjected to
                     dehydration-treated(1,2,5,10,24 hr)"
ORIGIN

Alignment Scores:
Pred. No.:              0.116          Length:        592
Score:                  84.00          Matches:       12
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     3              Gaps:          0

US-09-858-332G-15 (1-12) x BP796399 (1-592)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        114 CACAATCACAATCACAATCACAATCACAATCACAAT 79

RESULT 46
BJ429471/c
LOCUS       BJ429471                599 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ429471 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv3i14 3', mRNA sequence.
ACCESSION   BJ429471
VERSION     BJ429471.1  GI:19404193
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 599)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .599
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv3i14"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:              0.117          Length:        599
Score:                  84.00          Matches:       12
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     3              Gaps:          0

US-09-858-332G-15 (1-12) x BJ429471 (1-599)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        198 CACAATCACAATCACAATCACAATCACAATCACAAT 163

RESULT 47
BJ433687/c
LOCUS       BJ433687                604 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ433687 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv22113 3', mRNA sequence.
ACCESSION   BJ433687
VERSION     BJ433687.1  GI:19408409
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 604)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .604
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
```

```
                     /db_xref="taxon:44689"
                     /clone="ddv22113"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:                0.118          Length:         604
Score:                    84.00          Matches:        12
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       3              Gaps:           0

US-09-858-332G-15 (1-12) x BJ433687 (1-604)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        158 CACAATCACAATCACAATCACAATCACAATCACAAT 123

RESULT 48
BJ429190/c
LOCUS       BJ429190                 606 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ429190 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv2l16 3', mRNA sequence.
ACCESSION   BJ429190
VERSION     BJ429190.1  GI:19403912
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 606)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .606
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv2l16"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:                0.118          Length:         606
Score:                    84.00          Matches:        12
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       3              Gaps:           0

US-09-858-332G-15 (1-12) x BJ429190 (1-606)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        199 CACAATCACAATCACAATCACAATCACAATCACAAT 164

RESULT 49
BF237103
LOCUS       BF237103                 609 bp    mRNA    linear   EST 14-NOV-2000
DEFINITION  602027447F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:4162763 5',
            mRNA sequence.
ACCESSION   BF237103
VERSION     BF237103.1  GI:11151020
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 609)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9445  row: n  column: 12
            High quality sequence stop: 609.
FEATURES             Location/Qualifiers
     source          1. .609
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4162763"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Li9"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.9 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
ORIGIN

Alignment Scores:
Pred. No.:                0.119          Length:         609
Score:                    84.00          Matches:        12
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       2              Gaps:           0

US-09-858-332G-15 (1-12) x BF237103 (1-609)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db         83 CACAACCACAACCACAACCACAACCACAACCACAAC 118

RESULT 50
BJ433986/c
LOCUS       BJ433986                 609 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ433986 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv23e18 3', mRNA sequence.
ACCESSION   BJ433986
VERSION     BJ433986.1  GI:19408708
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 609)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
```

```
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES              Location/Qualifiers
     source           1. .609
                      /organism="Dictyostelium discoideum"
                      /mol_type="mRNA"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /clone="ddv23e18"
                      /sex="mat A"
                      /dev_stage="Growth phase"
                      /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:              0.119          Length:         609
Score:                  84.00          Matches:        12
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     3              Gaps:           0

US-09-858-332G-15 (1-12) x BJ433986 (1-609)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        152 CACAATCACAATCACAATCACAATCACAATCACAAT 117

RESULT 51
BQ783159/c
LOCUS       BQ783159                609 bp    mRNA    linear   EST 26-JUL-2002
DEFINITION  UI-R-FF0-cpi-d-10-0-UI.s1 NCI_CGAP_FF0 Rattus norvegicus cDNA clone
            UI-R-FF0-cpi-d-10-0-UI 3', mRNA sequence.
ACCESSION   BQ783159
VERSION     BQ783159.1  GI:21991631
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 609)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
   PUBMED   8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road , 4156  MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            Tissue Procurement: Jeff Stevens
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: DISTRIBUTION: Researchers may obtain clones
            from Research Genetics (www.resgen.com).
             The following repetitive elements were found in this cDNA
            sequence: 504-556, >(CAA)n#Simple_repeat (matched compliment)
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES              Location/Qualifiers
     source           1. .609
                      /organism="Rattus norvegicus"
                      /mol_type="mRNA"
                      /db_xref="taxon:10116"
                      /clone="UI-R-FF0-cpi-d-10-0-UI"
                      /tissue_type="Mixed tissues"
                      /dev_stage="Adult"
                      /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                      /clone_lib="NCI_CGAP_FF0"
                      /note="Vector: pT7T3-Pac (Pharmacia) with a modified
                      polylinker; Site_1: EcoR I; Site_2: Not I; UI-R-FF0 is a
                      subtracted cDNA library containing the following
                      tissue(s): Normal cartilage and SR-JWS Tumor Line . The
                      subtraction was made according to Bonaldo, Lennon and
                      Soares, Genome Research, 6:791-806, 1996. The
                      oligonucleotide used to prime the synthesis of
                      first-strand cDNA contains a library tag sequence that is
                      located between the Not I site and the (dT)18 tail. The
                      sequence tags for these libraries are: CTAATGGACG,
                      CATTCTTGTA.
                      TAG_TISSUE=rat SRC-JWST tumor line
                      TAG_LIB=UI-R-FF0
                      TAG_SEQ=CATTCTTGTA"
ORIGIN

Alignment Scores:
Pred. No.:              0.119          Length:         609
Score:                  84.00          Matches:        12
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     5              Gaps:           0

US-09-858-332G-15 (1-12) x BQ783159 (1-609)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        544 CACAACCATAACCACAACCACAACCACAACCACAAC 509

RESULT 52
BJ432773/c
LOCUS       BJ432773                611 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ432773 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv19c20 3', mRNA sequence.
ACCESSION   BJ432773
VERSION     BJ432773.1  GI:19407495
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 611)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES              Location/Qualifiers
     source           1. .611
                      /organism="Dictyostelium discoideum"
                      /mol_type="mRNA"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /clone="ddv19c20"
                      /sex="mat A"
                      /dev_stage="Growth phase"
                      /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:              0.119          Length:         611
Score:                  84.00          Matches:        12
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
```

```
Query Match:            100.00%       Indels:        0
DB:                     3             Gaps:          0

US-09-858-332G-15 (1-12) x BJ432773 (1-611)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        163 CACAATCACAATCACAATCACAATCACAATCACAAT 128

RESULT 53
BJ414802
LOCUS       BJ414802                616 bp    mRNA    linear   EST 10-MAR-2002
DEFINITION  BJ414802 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv20p07 5', mRNA sequence.
ACCESSION   BJ414802
VERSION     BJ414802.1  GI:19327313
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 616)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES              Location/Qualifiers
     source           1. .616
                      /organism="Dictyostelium discoideum"
                      /mol_type="mRNA"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /clone="ddv20p07"
                      /sex="mat A"
                      /dev_stage="Growth phase"
                      /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:              0.12          Length:        616
Score:                  84.00         Matches:       12
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     3             Gaps:          0

US-09-858-332G-15 (1-12) x BJ414802 (1-616)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        515 CACAACCACAACCACAACCACAACCACAACCACAAC 550

RESULT 54
BJ428216/c
LOCUS       BJ428216                616 bp    mRNA    linear   EST 11-MAR-2002
DEFINITION  BJ428216 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv11l01 3', mRNA sequence.
ACCESSION   BJ428216
VERSION     BJ428216.1  GI:19345573
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 616)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES              Location/Qualifiers
     source           1. .616
                      /organism="Dictyostelium discoideum"
                      /mol_type="mRNA"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /clone="ddv11l01"
                      /sex="mat A"
                      /dev_stage="Growth phase"
                      /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:              0.12          Length:        616
Score:                  84.00         Matches:       12
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     3             Gaps:          0

US-09-858-332G-15 (1-12) x BJ428216 (1-616)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        207 CACAATCACAATCACAATCACAATCACAATCACAAT 172

RESULT 55
BJ433790/c
LOCUS       BJ433790                618 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ433790 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv22n24 3', mRNA sequence.
ACCESSION   BJ433790
VERSION     BJ433790.1  GI:19408512
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 618)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES              Location/Qualifiers
     source           1. .618
                      /organism="Dictyostelium discoideum"
                      /mol_type="mRNA"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /clone="ddv22n24"
                      /sex="mat A"
                      /dev_stage="Growth phase"
                      /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:              0.121         Length:        618
Score:                  84.00         Matches:       12
Percent Similarity:     100.00%       Conservative:  0
```

```
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     3              Gaps:           0

US-09-858-332G-15 (1-12) x BJ433790 (1-618)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        158 CACAATCACAATCACAATCACAATCACAATCACAAT 123

RESULT 56
BJ431653/c
LOCUS       BJ431653                620 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ431653 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv15b01 3', mRNA sequence.
ACCESSION   BJ431653
VERSION     BJ431653.1  GI:19406375
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 620)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .620
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv15b01"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:              0.121          Length:         620
Score:                  84.00          Matches:        12
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     3              Gaps:           0

US-09-858-332G-15 (1-12) x BJ431653 (1-620)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        208 CACAATCACAATCACAATCACAATCACAATCACAAT 173

RESULT 57
BJ401757/c
LOCUS       BJ401757                622 bp    mRNA    linear   EST 10-MAR-2002
DEFINITION  BJ401757 Dictyostelium discoideum cDNA library, SF Dictyostelium
            discoideum cDNA clone dds19e12 3', mRNA sequence.
ACCESSION   BJ401757
VERSION     BJ401757.1  GI:19314674
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 622)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the slug stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .622
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dds19e12"
                     /sex="mat A"
                     /dev_stage="Slug stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, SF"
ORIGIN

Alignment Scores:
Pred. No.:              0.122          Length:         622
Score:                  84.00          Matches:        12
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     3              Gaps:           0

US-09-858-332G-15 (1-12) x BJ401757 (1-622)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        158 CACAATCACAATCACAATCACAATCACAATCACAAT 123

RESULT 58
BJ433898/c
LOCUS       BJ433898                622 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ433898 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv23d08 3', mRNA sequence.
ACCESSION   BJ433898
VERSION     BJ433898.1  GI:19408620
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 622)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .622
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv23d08"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:              0.122          Length:         622
Score:                  84.00          Matches:        12
Percent Similarity:     100.00%        Conservative:   0
```

```
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     3              Gaps:           0

US-09-858-332G-15 (1-12) x BJ433898 (1-622)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        163 CACAATCACAATCACAATCACAATCACAATCACAAT 128

RESULT 59
BJ340817/c
LOCUS       BJ340817                624 bp    mRNA    linear   EST 07-MAR-2002
DEFINITION  BJ340817 Dictyostelium discoideum cDNA library, AF Dictyostelium
            discoideum cDNA clone dda4c02 3', mRNA sequence.
ACCESSION   BJ340817
VERSION     BJ340817.1  GI:19249179
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 624)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the aggregation
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .624
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dda4c02"
                     /sex="mat A"
                     /dev_stage="Aggregation stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN

Alignment Scores:
Pred. No.:              0.122          Length:         624
Score:                  84.00          Matches:        12
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     3              Gaps:           0

US-09-858-332G-15 (1-12) x BJ340817 (1-624)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        167 CACAATCACAATCACAATCACAATCACAATCACAAT 132

RESULT 60
BM778956
LOCUS       BM778956                625 bp    mRNA    linear   EST 04-MAR-2002
DEFINITION  EST589531 KV2 Medicago truncatula cDNA clone pKV2-20C24, mRNA
            sequence.
ACCESSION   BM778956
VERSION     BM778956.1  GI:19108624
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 625)
  AUTHORS   VandenBosch,K., Endre,G., Hur,J., Beremand,P., Town,C.D., Van
            Aken,S., Utterback,T., Cheung,F., Tsai,J. and Fraser,C.M.
  TITLE     ESTs from roots of Medicago truncatula 48 hr after inoculation with
            Sinorhizobium meliloti
  JOURNAL   Unpublished (2002)
COMMENT     Other_ESTs: EST589530
            Contact: VandenBosch K
            Department of Plant Biology
            University of Minnesota
            220 BioSci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA
            Tel: 612 624 2755
            Fax: 612 625 1738
            Email: kvandenb@cbs.umn.edu
            TIGR sequence name: MTABH24TKB More information is available at:
            www.medicago.org
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES             Location/Qualifiers
     source          1. .625
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /cultivar="genotype A17"
                     /db_xref="taxon:3880"
                     /clone="pKV2-20C24"
                     /tissue_type="Seedling roots"
                     /dev_stage="2 days post-inoculation with Sinorhizobium
                     meliloti"
                     /lab_host="E. coli strain SOLR"
                     /clone_lib="KV2"
                     /note="Vector: pBluescript SK -; Site_1: EcoRI; Site_2:
                     XhoI; cDNA was prepared from polyA+ enriched RNA.  The
                     cDNA was  directionally ligated into the Unizap XR vector
                     from Stratagene  and packaged using Gigapack III Gold
                     packaging extracts.  Plasmids containing cDNA inserts
                     were excised from the  recombinant lambda-Zap phage using
                     Ex-assist helper phage and  propagated in SOLR cells."
ORIGIN

Alignment Scores:
Pred. No.:              0.122          Length:         625
Score:                  84.00          Matches:        12
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     3              Gaps:           0

US-09-858-332G-15 (1-12) x BM778956 (1-625)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        236 CACAATCATAATCATAATCATAATCACAATCACAAT 271

RESULT 61
BJ428793/c
LOCUS       BJ428793                626 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ428793 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv1c09 3', mRNA sequence.
ACCESSION   BJ428793
VERSION     BJ428793.1  GI:19403515
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 626)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
```

```
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES              Location/Qualifiers
     source           1. .626
                      /organism="Dictyostelium discoideum"
                      /mol_type="mRNA"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /clone="ddv1c09"
                      /sex="mat A"
                      /dev_stage="Growth phase"
                      /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:                0.123          Length:          626
Score:                    84.00          Matches:         12
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       3              Gaps:            0

US-09-858-332G-15 (1-12) x BJ428793 (1-626)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        170 CACAATCACAATCACAATCACAATCACAATCACAAT 135

RESULT 62
BJ432833/c
LOCUS       BJ432833                639 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ432833 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv20b02 3', mRNA sequence.
ACCESSION   BJ432833
VERSION     BJ432833.1  GI:19407555
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 639)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES              Location/Qualifiers
     source           1. .639
                      /organism="Dictyostelium discoideum"
                      /mol_type="mRNA"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /clone="ddv20b02"
                      /sex="mat A"
                      /dev_stage="Growth phase"
                      /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:                0.125          Length:          639
Score:                    84.00          Matches:         12
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       3              Gaps:            0

US-09-858-332G-15 (1-12) x BJ432833 (1-639)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        157 CACAATCACAATCACAATCACAATCACAATCACAAT 122

RESULT 63
CR320264
LOCUS       CR320264                644 bp    DNA     linear   GSS 19-NOV-2004
DEFINITION  mte1-46E22RM1 BAC end, cultivar Jemalong A17 of Medicago
            truncatula, genomic survey sequence.
ACCESSION   CR320264
VERSION     CR320264.1  GI:44866408
KEYWORDS    GSS.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 644)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
FEATURES              Location/Qualifiers
     source           1. .644
                      /organism="Medicago truncatula"
                      /mol_type="genomic DNA"
                      /cultivar="Jemalong A17"
                      /db_xref="taxon:3880"
                      /clone_lib="MTE1"
                      /note="Vector: pIndigoBAC ; Site_1: EcoRI ; Site_2: EcoRI
                      ; Debelle F. and Chalhoub B.
                      Genoscope sequence ID : mte1-46E22RM1"
ORIGIN

Alignment Scores:
Pred. No.:                0.126          Length:          644
Score:                    84.00          Matches:         12
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       11             Gaps:            0

US-09-858-332G-15 (1-12) x CR320264 (1-644)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        180 CACAATCATAATCATAATCATAATCACAATCACAAT 215

RESULT 64
BI271416
LOCUS       BI271416                652 bp    mRNA    linear   EST 18-JUL-2001
DEFINITION  NF051D09FL1F1078 Developing flower Medicago truncatula cDNA clone
            NF051D09FL 5', mRNA sequence.
ACCESSION   BI271416
VERSION     BI271416.1  GI:14879906
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 652)
  AUTHORS   Torres-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,
            Flores,H.R., Inman,J.T., Weller,J.W. and May,G.D.
  TITLE     Expressed Sequence Tags from the Samuel Roberts Noble Foundation
            Medicago truncatula flower library
  JOURNAL   Unpublished (2001)
COMMENT     Contact: May GD
            Plant Biology Division
            The Samuel Roberts Noble Foundation
```

```
           2510 Sam Noble Parkway, Ardmore, OK 73402, USA
           Tel: 580 224 6650
           Fax: 580 224 6692
           Email: gdmay@noble.org
           Insert Length: 652   Std Error: 0.00
           Plate: 051  row: D  column: 09
           Seq primer: TCACACAGGAAACAGCTATGAC.
FEATURES             Location/Qualifiers
     source          1. .652
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /db_xref="taxon:3880"
                     /clone="NF051D09FL"
                     /tissue_type="Developing flowers"
                     /dev_stage="Developmentally pooled. Contains a mixture of
                     very young, developing, fully-opened flowers and flowers
                     in early transition into pods."
                     /clone_lib="Developing flower"
                     /note="Vector: Lambda Zap; cDNA was prepared from polyA+
                     enriched, pooled samples of equivalent  amounts of total
                     RNA from very young, developing, fully-opened flowers  and
                     flowers transitioning into pods. The cDNA was
                     directionally ligated  into the Uni-Zap XR vector
                     (Stratagene) and packaged using the Gigapack  III Gold
                     packaging extracts. Phagemids containing cDNA inserts were
                     in  vivo excised from the recombinant Uni-ZAP XR vector
                     using ExAssist helper  phage and the E. coli strain
                     XL1-Blue MRF' (Stratagene). Excised  plasmids were plated
                     using SOLR cells."
ORIGIN

Alignment Scores:
Pred. No.:                0.128          Length:         652
Score:                    84.00          Matches:        12
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       2              Gaps:           0

US-09-858-332G-15 (1-12) x BI271416 (1-652)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        426 CACAATCATAATCATAATCATAATCACAATCACAAT 461

RESULT 65
BJ432931/c
LOCUS       BJ432931                652 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ432931 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv20d08 3', mRNA sequence.
ACCESSION   BJ432931
VERSION     BJ432931.1  GI:19407653
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 652)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .652
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv20d08"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:                0.128          Length:         652
Score:                    84.00          Matches:        12
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       3              Gaps:           0

US-09-858-332G-15 (1-12) x BJ432931 (1-652)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        171 CACAATCACAATCACAATCACAATCACAATCACAAT 136

RESULT 66
BJ431681/c
LOCUS       BJ431681                657 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ431681 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv15h02 3', mRNA sequence.
ACCESSION   BJ431681
VERSION     BJ431681.1  GI:19406403
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 657)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .657
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv15h02"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:                0.129          Length:         657
Score:                    84.00          Matches:        12
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       3              Gaps:           0

US-09-858-332G-15 (1-12) x BJ431681 (1-657)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        209 CACAATCACAATCACAATCACAATCACAATCACAAT 174

RESULT 67
CD373395
LOCUS       CD373395                658 bp    mRNA    linear   EST 29-MAY-2003
```

```
DEFINITION  UI-R-GR0-csv-n-10-0-UI.r1 UI-R-GR0 Rattus norvegicus cDNA clone
            UI-R-GR0-csv-n-10-0-UI 5', mRNA sequence.
ACCESSION   CD373395
VERSION     CD373395.1  GI:31157485
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 658)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
   PUBMED   8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road , 4156  MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            Tissue Procurement: James Lin, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Distribution information can be found at
            http://genome.uiowa.edu/distribution/rat.html
            Seq primer: M13 REVERSE.
FEATURES             Location/Qualifiers
     source          1. .658
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-GR0-csv-n-10-0-UI"
                     /tissue_type="Whole embryo"
                     /dev_stage="embryo 13dpc"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /clone_lib="UI-R-GR0"
                     /note="Vector: pYX-Asc; Site_1: EcoR I; Site_2: Not I;
                     UI-R-GR0 is a cDNA library containing the following
                     tissue(s): rat whole embryo 13dpc. The library was
                     constructed according to Bonaldo, Lennon and Soares,
                     Genome Research, 6:791-806, 1996. Denatured RNA was size
                     fractionated on a 1% agarose gel. First strand cDNA
                     synthesis was primed with oligo-dT primer containing a Not
                     I site. Double strand cDNA was size selected according to
                     mRNA size fraction, ligated with EcoR I adaptor, digested
                     with NotI and then cloned directionally into pYX-Asc
                     vector. The library tag sequence located between the Not I
                     site and the polyA tail is CATCTCTACT. This library was
                     created for the University of Iowa Program for Rat Gene
                     Discovery and Mapping (Val Sheffield, Bento Soares and Tom
                     Casavant)."
ORIGIN

Alignment Scores:
Pred. No.:              0.129          Length:         658
Score:                  84.00          Matches:        12
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     6              Gaps:           0

US-09-858-332G-15 (1-12) x CD373395 (1-658)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db         99 CACAACCATAACCACAACCACAACCACAACCACAAC 134

RESULT 68
```

```
BQ781099/c
LOCUS       BQ781099                661 bp    mRNA    linear   EST 26-JUL-2002
DEFINITION  UI-R-FF0-cpd-j-13-0-UI.s1 NCI_CGAP_FF0 Rattus norvegicus cDNA clone
            UI-R-FF0-cpd-j-13-0-UI 3', mRNA sequence.
ACCESSION   BQ781099
VERSION     BQ781099.1  GI:21989571
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 661)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
   PUBMED   8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road , 4156  MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            Tissue Procurement: Jeff Stevens
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: DISTRIBUTION: Researchers may obtain clones
            from Research Genetics (www.resgen.com).
             The following repetitive elements were found in this cDNA
            sequence: 504-556, >(CAA)n#Simple_repeat (matched compliment)
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .661
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
                     /clone="UI-R-FF0-cpd-j-13-0-UI"
                     /tissue_type="Mixed tissues"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /clone_lib="NCI_CGAP_FF0"
                     /note="Vector: pT7T3-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoR I; Site_2: Not I; UI-R-FF0 is a
                     subtracted cDNA library containing the following
                     tissue(s): Normal cartilage and SR-JWS Tumor Line . The
                     subtraction was made according to Bonaldo, Lennon and
                     Soares, Genome Research, 6:791-806, 1996. The
                     oligonucleotide used to prime the synthesis of
                     first-strand cDNA contains a library tag sequence that is
                     located between the Not I site and the (dT)18 tail. The
                     sequence tags for these libraries are: CTAATGGACG,
                     CATTCTTGTA.
                     TAG_TISSUE=cartilage
                     TAG_LIB=UI-R-FF0
                     TAG_SEQ=CTAATGGACG"
ORIGIN

Alignment Scores:
Pred. No.:              0.13           Length:         661
Score:                  84.00          Matches:        12
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     5              Gaps:           0

US-09-858-332G-15 (1-12) x BQ781099 (1-661)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
```

```
Db        544 CACAACCATAACCACAACCACAACCACAACCACAAC 509

RESULT 69
BI281721/c
LOCUS       BI281721                662 bp    mRNA    linear   EST 19-JUL-2001
DEFINITION  UI-R-CT0s-cav-f-10-0-UI.s1 UI-R-CT0s Rattus norvegicus cDNA clone
            UI-R-CT0s-cav-f-10-0-UI 3', mRNA sequence.
ACCESSION   BI281721
VERSION     BI281721.1  GI:14931752
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 662)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
   PUBMED   8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road , 4156  MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            non-normalized rat brain pool library cDNA Library Preparation:
            M.B. Soares Lab Clone distribution: clones will be available
            through Research Genetics (www.resgen.com) The following repetitive
            elements were found in this cDNA sequence: 504-556,
            >(CAA)n#Simple_repeat
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES              Location/Qualifiers
     source           1. .662
                      /organism="Rattus norvegicus"
                      /mol_type="mRNA"
                      /strain="Sprague-Dawley"
                      /db_xref="taxon:10116"
                      /clone="UI-R-CT0s-cav-f-10-0-UI"
                      /dev_stage="ADULT"
                      /lab_host="DH10B (Life Technologies)"
                      /clone_lib="UI-R-CT0s"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CT0s
                      library is a non-normalized library constructed from the
                      following rat brain tissues: embryonic day 17, embryonic
                      day 19, embryonic day 21, adult day 1, adult day 12, adult
                      day 75, adult day 200. For a detailed description of the
                      library from which this clone was derived, please visit
                      our web site at ratest.eng.uiowa.edu. The subtraction has
                      been previously described in (Bonaldo, Lennon and Soares,
                      Genome Research 6:791-806, 1996)
                      TAG_TISSUE=rat brain pool
                      TAG_LIB=UI-R-CT0s
                      TAG_SEQ=ACTTC"
ORIGIN

Alignment Scores:
Pred. No.:              0.13          Length:         662
Score:                  84.00         Matches:        12
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     2             Gaps:           0

US-09-858-332G-15 (1-12) x BI281721 (1-662)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        544 CACAACCATAACCACAACCACAACCACAACCACAAC 509

RESULT 70
AV823854/c
LOCUS       AV823854                663 bp    mRNA    linear   EST 01-APR-2002
DEFINITION  AV823854 RAFL6 Arabidopsis thaliana cDNA clone RAFL06-09-I21 5',
            mRNA sequence.
ACCESSION   AV823854
VERSION     AV823854.1  GI:19865914
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 663)
  AUTHORS   Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
            Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,
            Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y.
            and Shinozaki,K.
  TITLE     Large scale analysis of Arabidopsis full-length cDNA (2002b)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            An Arabidopsis full-length cDNA library was constructed essentially
            as reported previously (Seki et al., 1998).This clone is in a
            modified pBluescript vector as a SstI/XhoI insert. Please visit our
            web site (http://www.gsc.riken.go.jp/e/plant/index _e.html) for
            further details.
FEATURES              Location/Qualifiers
     source           1. .663
                      /organism="Arabidopsis thaliana"
                      /mol_type="mRNA"
                      /db_xref="taxon:3702"
                      /clone="RAFL06-09-I21"
                      /dev_stage="plants at various developmental stages from
                      germination to mature seeds"
                      /lab_host="DH10B"
                      /clone_lib="RAFL6"
                      /note="Site_1: SstI; Site_2: XhoI; subjected to
                      dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24
                      hr) treatments"
ORIGIN

Alignment Scores:
Pred. No.:              0.13          Length:         663
Score:                  84.00         Matches:        12
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     1             Gaps:           0

US-09-858-332G-15 (1-12) x AV823854 (1-663)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        144 CACAATCACAATCACAATCACAATCACAATCACAAT 109

RESULT 71
BJ429328/c
LOCUS       BJ429328                663 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ429328 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv3i04 3', mRNA sequence.
ACCESSION   BJ429328
```

```
VERSION     BJ429328.1  GI:19404050
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 663)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .663
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv3i04"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:                0.13          Length:          663
Score:                    84.00         Matches:         12
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       3             Gaps:            0

US-09-858-332G-15 (1-12) x BJ429328 (1-663)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        198 CACAATCACAATCACAATCACAATCACAATCACAAT 163

RESULT 72
BJ433069/c
LOCUS       BJ433069                667 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ433069 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv20n16 3', mRNA sequence.
ACCESSION   BJ433069
VERSION     BJ433069.1  GI:19407791
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 667)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .667
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv20n16"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:                0.131         Length:          667
Score:                    84.00         Matches:         12
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       3             Gaps:            0

US-09-858-332G-15 (1-12) x BJ433069 (1-667)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        146 CACAATCACAATCACAATCACAATCACAATCACAAT 111

RESULT 73
BJ430907/c
LOCUS       BJ430907                668 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ430907 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv9a09 3', mRNA sequence.
ACCESSION   BJ430907
VERSION     BJ430907.1  GI:19405629
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 668)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .668
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv9a09"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:                0.131         Length:          668
Score:                    84.00         Matches:         12
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       3             Gaps:            0

US-09-858-332G-15 (1-12) x BJ430907 (1-668)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        200 CACAATCACAATCACAATCACAATCACAATCACAAT 165

RESULT 74
BJ434653/c
LOCUS       BJ434653                668 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ434653 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv24f16 3', mRNA sequence.
```

```
ACCESSION   BJ434653
VERSION     BJ434653.1  GI:19409375
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 668)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .668
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv24f16"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:              0.131          Length:         668
Score:                  84.00          Matches:        12
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     3              Gaps:           0

US-09-858-332G-15 (1-12) x BJ434653 (1-668)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        152 CACAATCACAATCACAATCACAATCACAATCACAAT 117

RESULT 75
BJ431437/c
LOCUS       BJ431437                676 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ431437 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv14m03 3', mRNA sequence.
ACCESSION   BJ431437
VERSION     BJ431437.1  GI:19406159
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 676)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .676
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv14m03"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:              0.133          Length:         676
Score:                  84.00          Matches:        12
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     3              Gaps:           0

US-09-858-332G-15 (1-12) x BJ431437 (1-676)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        194 CACAATCACAATCACAATCACAATCACAATCACAAT 159

RESULT 76
CV521375
LOCUS       CV521375                676 bp    mRNA    linear   EST 06-OCT-2004
DEFINITION  0089P0058Z.x0_G04 Mimulus guttatus library 2 Mimulus guttatus cDNA
            clone 0089P0058Z.x0_G04, mRNA sequence.
ACCESSION   CV521375
VERSION     CV521375.1  GI:53847907
KEYWORDS    EST.
SOURCE      Mimulus guttatus (spotted monkey flower)
  ORGANISM  Mimulus guttatus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Lamiales; Phrymaceae; Mimulus.
REFERENCE   1  (bases 1 to 676)
  AUTHORS   Willis,J., Vision,T., Dietrich,F.S. and Allen,A.
  TITLE     Mimulus guttatus cDNA sequence
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Willis J
            Department of Biology
            Duke University
            072-A Biological Sciences Science Drive, Durham, NC 27708, USA
            Tel: 919 660 7340
            Fax: 919 660 7293
            Email: jwillis@duke.edu
            Plate: 0089P0058  row: 04  column: G
            Seq primer: T7
            High quality sequence start: 16
            High quality sequence stop: 704.
FEATURES             Location/Qualifiers
     source          1. .676
                     /organism="Mimulus guttatus"
                     /mol_type="mRNA"
                     /db_xref="taxon:4155"
                     /clone="0089P0058Z.x0_G04"
                     /clone_lib="Mimulus guttatus library 2"
                     /note="Vector: pGEM-T Easy; a Mimulus guttatus cDNA
                     library"
ORIGIN

Alignment Scores:
Pred. No.:              0.133          Length:         676
Score:                  84.00          Matches:        12
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     7              Gaps:           0

US-09-858-332G-15 (1-12) x CV521375 (1-676)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        432 CATAATCATAATCATAATCACAATCACAATCACAAT 467
```

```
RESULT 77
BJ429677/c
LOCUS       BJ429677                677 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ429677 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv4c17 3', mRNA sequence.
ACCESSION   BJ429677
VERSION     BJ429677.1  GI:19404399
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 677)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .677
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv4c17"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:                0.133          Length:          677
Score:                    84.00          Matches:         12
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       3              Gaps:            0

US-09-858-332G-15 (1-12) x BJ429677 (1-677)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        200 CACAATCACAATCACAATCACAATCACAATCACAAT 165

RESULT 78
BG448016
LOCUS       BG448016                678 bp    mRNA    linear   EST 16-MAR-2001
DEFINITION  NF104H02EC1F1026 Elicited cell culture Medicago truncatula cDNA
            clone NF104H02EC 5', mRNA sequence.
ACCESSION   BG448016
VERSION     BG448016.1  GI:13366797
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 678)
  AUTHORS   Torres-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,
            Flores,H.R., Inman,J.T., Weller,J.W. and May,G.D.
  TITLE     Expressed Sequence Tags from the Samuel Roberts Noble Foundation -
            Center for Medicago Genomics Research
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Dixon RA
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA
            Tel: 580 221 7302
            Fax: 580 221 7380
            Email: radixon@noble.org
            Insert Length: 678   Std Error: 0.00
            Plate: 104  row: H  column: 02
            Seq primer: TCACACAGGAAACAGCTATGAC.
FEATURES             Location/Qualifiers
     source          1. .678
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /db_xref="taxon:3880"
                     /clone="NF104H02EC"
                     /tissue_type="Cell cultures derived from root tissues"
                     /dev_stage="Cell suspensions were subcultured every 14
                     days. Cells were induced six days after subculture"
                     /clone_lib="Elicited cell culture"
                     /note="Vector: Lambda Zap; Cells were induced with yeast
                     cell wall extracts equivalent to  50ug/ml glucose in the
                     final concentration. Samples were taken at  0.5, 1, 12 and
                     24 hours after induction. Equal amounts of RNA from  each
                     time point were pooled and used for mRNA isolation."
ORIGIN

Alignment Scores:
Pred. No.:                0.133          Length:          678
Score:                    84.00          Matches:         12
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       2              Gaps:            0

US-09-858-332G-15 (1-12) x BG448016 (1-678)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        264 CACAATCATAATCATAATCATAATCACAATCACAAT 299

RESULT 79
BJ431232/c
LOCUS       BJ431232                679 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ431232 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv13h09 3', mRNA sequence.
ACCESSION   BJ431232
VERSION     BJ431232.1  GI:19405954
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 679)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .679
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv13h09"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
```

```
Pred. No.:              0.133          Length:        679
Score:                  84.00          Matches:       12
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     3              Gaps:          0

US-09-858-332G-15 (1-12) x BJ431232 (1-679)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        203 CACAATCACAATCACAATCACAATCACAATCACAAT 168

RESULT 80
BJ430517/c
LOCUS       BJ430517                690 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ430517 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv7d17 3', mRNA sequence.
ACCESSION   BJ430517
VERSION     BJ430517.1  GI:19405239
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 690)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .690
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv7d17"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:              0.136          Length:        690
Score:                  84.00          Matches:       12
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     3              Gaps:          0

US-09-858-332G-15 (1-12) x BJ430517 (1-690)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        118 CACAATCACAATCACAATCACAATCACAATCACAAT 83

RESULT 81
CG931836
LOCUS       CG931836                695 bp    DNA     linear   GSS 12-DEC-2003
DEFINITION  MBEAL86TR mth2 Medicago truncatula genomic clone 13P4, genomic
            survey sequence.
ACCESSION   CG931836
VERSION     CG931836.1  GI:39797669
KEYWORDS    GSS.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 695)
  AUTHORS   Town,C.D., Shetty,J., Koo,H. and Feldblyum,T.F.
  TITLE     Sequencing of BAC ends from Medicago truncatula
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: MBEAL86TFB
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            Seq primer: CAGGAAACAGCTATGACC
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .695
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /cultivar="genotype A17"
                     /db_xref="taxon:3880"
                     /clone="13P4"
                     /clone_lib="mth2"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII; Cook, D.R. and Kim, D.J, unpublished"
ORIGIN

Alignment Scores:
Pred. No.:              0.137          Length:        695
Score:                  84.00          Matches:       12
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     10             Gaps:          0

US-09-858-332G-15 (1-12) x CG931836 (1-695)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        177 CACAATCATAATCATAATCATAATCACAATCACAAT 212

RESULT 82
BJ434010/c
LOCUS       BJ434010                697 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ434010 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv23j14 3', mRNA sequence.
ACCESSION   BJ434010
VERSION     BJ434010.1  GI:19408732
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 697)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .697
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv23j14"
                     /sex="mat A"
```

```
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:                0.137          Length:          697
Score:                    84.00          Matches:         12
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       3              Gaps:            0

US-09-858-332G-15 (1-12) x BJ434010 (1-697)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        201 CACAATCACAATCACAATCACAATCACAATCACAAT 166

RESULT 83
CO389214
LOCUS       CO389214                 697 bp    mRNA    linear   EST 30-JUN-2004
DEFINITION  AGENCOURT_26623004 NIH_MGC_253 Rattus norvegicus cDNA clone
            IMAGE:7303816 5', mRNA sequence.
ACCESSION   CO389214
VERSION     CO389214.1  GI:49496805
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 697)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Drs. Josef Lazar & Howard Jacob, Medical
            College of Wisconsin
             cDNA Library Preparation: Express Genomics
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM15339  row: g  column: 14
            High quality sequence stop: 666.
FEATURES             Location/Qualifiers
     source          1. .697
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
                     /clone="IMAGE:7303816"
                     /tissue_type="Ovary - Brown Norway Line 3 Age 8 weeks.
                     Tissues were snap-frozen and transfered in -70c. RNAse
                     free the entire procedure."
                     /lab_host="DH10B TonA"
                     /clone_lib="NIH_MGC_253"
                     /note="Organ: ovary; Vector: pExpress-1; Site_1: EcoRV;
                     Site_2: NotI; RNA obtained from female animals at 8 wk
                     old. Tissues were snap-frozen and kept at -80C for two
                     days before RNA extraction and purification (Tri-reagent
                     method). cDNA was primed using oligo-dT primer:
                     5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC(T)25-3' and cloned into
                     the EcoRV/NotI sites of pExpress-1. Size-selection >1.25kb
                     resulted in an average insert size of 1.5 kb. This primary
                     library is normalized (non-normalized primary library is
                     NIH_MGC_252) and was constructed by Express Genomics
                     (Frederick, MD). Note: this is a NIH_MGC library."
ORIGIN

Alignment Scores:
Pred. No.:                0.137          Length:          697
Score:                    84.00          Matches:         12
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       7              Gaps:            0

US-09-858-332G-15 (1-12) x CO389214 (1-697)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        281 CACAACCATAACCACAACCACAACCACAACCACAAC 316

RESULT 84
BJ432382/c
LOCUS       BJ432382                 701 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ432382 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv18d16 3', mRNA sequence.
ACCESSION   BJ432382
VERSION     BJ432382.1  GI:19407104
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 701)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .701
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv18d16"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:                0.138          Length:          701
Score:                    84.00          Matches:         12
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       3              Gaps:            0

US-09-858-332G-15 (1-12) x BJ432382 (1-701)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        159 CACAATCACAATCACAATCACAATCACAATCACAAT 124

RESULT 85
AQ874103/c
LOCUS       AQ874103                 714 bp    DNA     linear   GSS 08-NOV-1999
DEFINITION  V103A5 mTn-3xHA/lacZ Insertion Library, strain Y2278 Saccharomyces
            cerevisiae genomic 5', genomic survey sequence.
ACCESSION   AQ874103
VERSION     AQ874103.1  GI:6286347
KEYWORDS    GSS.
SOURCE      Saccharomyces cerevisiae (baker's yeast)
```

```
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 714)
  AUTHORS   Ross-Macdonald,P., Roemer,T., Coelho,P.S.R., Agarwal,S., Kumar,A.,
            desEtages,S.A., Cheung,K.-H., Sheehan,A., Symoniatis,D., Jansen,R.,
            Umansky,L., Heidtman,M., Nelson,K., Iwasaki,H., Kanada,D., Lugo,R.,
            Hager,K., Miller,P., Roeder,G.S. and Snyder,M.
  TITLE     Large-Scale Analysis of the Yeast Genome by Transposon Tagging and
            Gene Disruption
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Kumar A
            Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
            Yale University
            P.O. Box 208103, New Haven, CT 06520-8103, USA
            Tel: 203 432 9949
            Fax: 203 432 6161
            Email: anuj.kumar@yale.edu
            te of mTn-3xHA/lacZ insertion.
            Seq primer: GGCCTTCTTTCTTTGGAAGTAC
            Class: transposon-tagged.
FEATURES             Location/Qualifiers
     source          1. .714
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /strain="Y2278 - S288C background, cir(0) rho(0)"
                     /db_xref="taxon:4932"
                     /lab_host="E. coli"
                     /clone_lib="mTn-3xHA/lacZ Insertion Library, strain Y2278"
                     /note="Vector: pHSS6-Sal; A yeast genomic DNA library
                     without 2 micron or mitochondrial DNA  was prepared in
                     pHSS6-Sal; genomic DNA was size-fractionated  (DNA of
                     roughly 2-3 kb in length) prior to cloning. This library
                     was subsequently mutagenized with a mTn-3xHA/lacZ
                     minitransposon  containing lacZ, URA3, and tet resistance.
                     "
ORIGIN

Alignment Scores:
Pred. No.:              0.141          Length:         714
Score:                  84.00          Matches:        12
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     9              Gaps:           0

US-09-858-332G-15 (1-12) x AQ874103 (1-714)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        337 CATAATCATAATCATAATCATAATCATAATCATAAT 302

RESULT 86
BJ430955/c
LOCUS       BJ430955                717 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ430955 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv9j12 3', mRNA sequence.
ACCESSION   BJ430955
VERSION     BJ430955.1  GI:19405677
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 717)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .717
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv9j12"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:              0.141          Length:         717
Score:                  84.00          Matches:        12
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     3              Gaps:           0

US-09-858-332G-15 (1-12) x BJ430955 (1-717)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        155 CACAATCACAATCACAATCACAATCACAATCACAAT 120

RESULT 87
BJ445670/c
LOCUS       BJ445670                718 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ445670 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv60b11 3', mRNA sequence.
ACCESSION   BJ445670
VERSION     BJ445670.1  GI:19420391
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 718)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .718
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv60b11"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:              0.141          Length:         718
Score:                  84.00          Matches:        12
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     3              Gaps:           0

US-09-858-332G-15 (1-12) x BJ445670 (1-718)
```

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        158 CACAATCACAATCACAATCACAATCACAATCACAAT 123

RESULT 88
BJ429204/c
LOCUS       BJ429204                723 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ429204 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv2o13 3', mRNA sequence.
ACCESSION   BJ429204
VERSION     BJ429204.1  GI:19403926
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 723)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .723
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv2o13"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:                0.142          Length:         723
Score:                    84.00          Matches:        12
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       3              Gaps:           0

US-09-858-332G-15 (1-12) x BJ429204 (1-723)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        200 CACAATCACAATCACAATCACAATCACAATCACAAT 165

RESULT 89
BJ434636/c
LOCUS       BJ434636                724 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ434636 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv24b15 3', mRNA sequence.
ACCESSION   BJ434636
VERSION     BJ434636.1  GI:19409358
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 724)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .724
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv24b15"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:                0.143          Length:         724
Score:                    84.00          Matches:        12
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       3              Gaps:           0

US-09-858-332G-15 (1-12) x BJ434636 (1-724)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        202 CACAATCACAATCACAATCACAATCACAATCACAAT 167

RESULT 90
BJ429573/c
LOCUS       BJ429573                725 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ429573 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv4c02 3', mRNA sequence.
ACCESSION   BJ429573
VERSION     BJ429573.1  GI:19404295
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 725)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .725
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv4c02"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:                0.143          Length:         725
Score:                    84.00          Matches:        12
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       3              Gaps:           0

US-09-858-332G-15 (1-12) x BJ429573 (1-725)
```

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        203 CACAATCACAATCACAATCACAATCACAATCACAAT 168

RESULT 91
CN834352
LOCUS       CN834352                725 bp    mRNA    linear   EST 02-JUN-2004
DEFINITION  AGENCOURT_24954489 NIH_MGC_169 Mus musculus cDNA clone
            IMAGE:30917704 5', mRNA sequence.
ACCESSION   CN834352
VERSION     CN834352.1  GI:47938924
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 725)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jonathan Kuo, NIMH
             cDNA Library Preparation: Michael Brownstein Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDCM384  row: f  column: 17
            High quality sequence stop: 619.
FEATURES             Location/Qualifiers
     source          1. .725
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:30917704"
                     /lab_host="DH10B (T1-phage-resistant)"
                     /clone_lib="NIH_MGC_169"
                     /note="Organ: Testicles; Vector: pDNR-LIB; Site_1: SfiI
                     (ggccattatggcc); Site_2: SfiI (ggccgcctcggcc); cDNA made
                     by oligo-dT priming and directionally cloned. 5' and 3'
                     adaptors were used in cloning as follows:
                     5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3' and
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)NN-3'. Full-length
                     enriched library was constructed using the Clontech
                     Creator SMART kit and size-selected to contain the 0.5 kb
                     size fraction. Library created in thelaboratory of M.
                     Brownstein (NIMH, NIH). Note: this is a NIH_MGC Library."
ORIGIN

Alignment Scores:
Pred. No.:                0.143          Length:          725
Score:                    84.00          Matches:         12
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       7              Gaps:            0

US-09-858-332G-15 (1-12) x CN834352 (1-725)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        106 CACAACCACAACCACAACCACAACCACAACCACAAC 141

RESULT 92
BJ436240/c
LOCUS       BJ436240                731 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ436240 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv30c16 3', mRNA sequence.
ACCESSION   BJ436240
VERSION     BJ436240.1  GI:19410962
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 731)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .731
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv30c16"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:                0.144          Length:          731
Score:                    84.00          Matches:         12
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       3              Gaps:            0

US-09-858-332G-15 (1-12) x BJ436240 (1-731)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        168 CACAATCACAATCACAATCACAATCACAATCACAAT 133

RESULT 93
CR277961
LOCUS       CR277961                735 bp    DNA     linear   GSS 06-JUL-2004
DEFINITION  Forward strand read from insert in 5'HPRT insertion targeting and
            chromosome engineering clone MHPN365o14, genomic survey sequence.
ACCESSION   CR277961
VERSION     CR277961.1  GI:50056938
KEYWORDS    GSS; genome survey sequence; MICER.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 735)
  AUTHORS   Adams,D.J., Biggs,P.J., Cox,A.V., Davies,R.M., van der Weyden,L.,
            Jonkers,J., Smith,J., Plumb,R.W., Taylor,R.G., Nishijima,I., Yu,Y.,
            Rogers,J. and Bradley,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. http://www.sanger.ac.uk/MICER
FEATURES             Location/Qualifiers
     source          1. .735
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="MHPN365o14"
                     /clone_lib="MHPN"
```

```
ORIGIN

Alignment Scores:
Pred. No.:                0.145          Length:         735
Score:                    84.00          Matches:        12
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       11             Gaps:           0

US-09-858-332G-15 (1-12) x CR277961 (1-735)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        441 CACAACCACAACCACAACCACAACCACAACCACAAC 476

RESULT 94
BF385371
LOCUS       BF385371                736 bp    mRNA    linear   EST 27-NOV-2000
DEFINITION  602047107F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:4196544 5',
            mRNA sequence.
ACCESSION   BF385371
VERSION     BF385371.1  GI:11366676
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 736)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9533  row: n  column: 01
            High quality sequence stop: 735.
FEATURES             Location/Qualifiers
     source          1. .736
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4196544"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Li9"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.9 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
ORIGIN

Alignment Scores:
Pred. No.:                0.145          Length:         736
Score:                    84.00          Matches:        12
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       2              Gaps:           0

US-09-858-332G-15 (1-12) x BF385371 (1-736)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        184 CACAACCACAACCACAACCACAACCACAACCACAAC 219

RESULT 95
C93043
LOCUS       C93043                  738 bp    mRNA    linear   EST 12-JUL-1999
DEFINITION  C93043 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
            discoideum cDNA clone SSF591, mRNA sequence.
ACCESSION   C93043
VERSION     C93043.1  GI:3074919
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 738)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     The Dictyostelium developmental cDNA project: generation and
            analysis of expressed sequence tags from the first-finger stage of
            development
  JOURNAL   DNA Res. 5 (6), 335-340 (1998)
  PUBMED    10048482
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp
            PROJECT = 'Dictyostelium discoideum cDNA project in Japan'
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .738
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SSF591"
                     /dev_stage="slug"
                     /clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
ORIGIN

Alignment Scores:
Pred. No.:                0.145          Length:         738
Score:                    84.00          Matches:        12
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       5              Gaps:           0

US-09-858-332G-15 (1-12) x C93043 (1-738)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        442 CACAATCACAATCACAATCACAATCACAATCACAAT 477

RESULT 96
CO387676
LOCUS       CO387676                738 bp    mRNA    linear   EST 30-JUN-2004
DEFINITION  AGENCOURT_26626266 NIH_MGC_253 Rattus norvegicus cDNA clone
            IMAGE:7301812 5', mRNA sequence.
ACCESSION   CO387676
VERSION     CO387676.1  GI:49495267
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 738)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
```

```
           National Cancer Institute / NIH
           Bldg. 31 Rm10A07 Bethesda, MD 20892
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: Drs. Josef Lazar & Howard Jacob, Medical
           College of Wisconsin
            cDNA Library Preparation: Express Genomics
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM15334  row: d  column: 02
           High quality sequence stop: 663.
FEATURES             Location/Qualifiers
     source          1. .738
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
                     /clone="IMAGE:7301812"
                     /tissue_type="Ovary - Brown Norway Line 3 Age 8 weeks.
                     Tissues were snap-frozen and transfered in -70c. RNAse
                     free the entire procedure."
                     /lab_host="DH10B TonA"
                     /clone_lib="NIH_MGC_253"
                     /note="Organ: ovary; Vector: pExpress-1; Site_1: EcoRV;
                     Site_2: NotI; RNA obtained from female animals at 8 wk
                     old. Tissues were snap-frozen and kept at -80C for two
                     days before RNA extraction and purification (Tri-reagent
                     method). cDNA was primed using oligo-dT primer:
                     5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC(T)25-3' and cloned into
                     the EcoRV/NotI sites of pExpress-1. Size-selection >1.25kb
                     resulted in an average insert size of 1.5 kb. This primary
                     library is normalized (non-normalized primary library is
                     NIH_MGC_252) and was constructed by Express Genomics
                     (Frederick, MD). Note: this is a NIH_MGC library."
ORIGIN

Alignment Scores:
Pred. No.:                0.145          Length:         738
Score:                    84.00          Matches:        12
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       7              Gaps:           0

US-09-858-332G-15 (1-12) x CO387676 (1-738)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        281 CACAACCATAACCACAACCACAACCACAACCACAAC 316

RESULT 97
BJ401045/c
LOCUS       BJ401045                741 bp    mRNA    linear   EST 10-MAR-2002
DEFINITION  BJ401045 Dictyostelium discoideum cDNA library, SF Dictyostelium
            discoideum cDNA clone dds15o24 3', mRNA sequence.
ACCESSION   BJ401045
VERSION     BJ401045.1  GI:19313962
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 741)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the slug stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .741
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dds15o24"
                     /sex="mat A"
                     /dev_stage="Slug stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, SF"
ORIGIN

Alignment Scores:
Pred. No.:                0.146          Length:         741
Score:                    84.00          Matches:        12
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       3              Gaps:           0

US-09-858-332G-15 (1-12) x BJ401045 (1-741)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        173 CACAATCACAATCACAATCACAATCACAATCACAAT 138

RESULT 98
BX211625/c
LOCUS       BX211625                749 bp    DNA     linear   GSS 29-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-250M24, genomic survey sequence.
ACCESSION   BX211625
VERSION     BX211625.1  GI:28043511
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 749)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 250M24. 250M24
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .749
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-250M24"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

Alignment Scores:
Pred. No.:                0.148          Length:         749
Score:                    84.00          Matches:        12
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       10             Gaps:           0

US-09-858-332G-15 (1-12) x BX211625 (1-749)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        485 CACAATCACAATCACAATCACAATCACAATCACAAT 450

RESULT 99
```

```
CZ281894/c
LOCUS       CZ281894                752 bp    DNA     linear   GSS 01-JUL-2005
DEFINITION  cp19e05.f Candida parapsilosis Random Genomic Library Candida
            parapsilosis genomic clone cp19e05, genomic survey sequence.
ACCESSION   CZ281894
VERSION     CZ281894.1  GI:68453096
KEYWORDS    GSS.
SOURCE      Candida parapsilosis
  ORGANISM  Candida parapsilosis
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 752)
  AUTHORS   Logue,M.E., Wong,S., Wolfe,K.H. and Butler,G.
  TITLE     A genome sequence survey shows that the pathogenic yeast Candida
            parapsilosis has a defective MTLa1 allele at its mating type locus
  JOURNAL   Eukaryot. Cell 4 (6), 1009-1017 (2005)
   PUBMED   15947193
COMMENT     Contact: Logue M
            Department of Biochemistry, Conway Institute of Biomolecular and
            Biomedical Research
            University College Dublin
            Dublin 4, Ireland
            Tel: +353 1 7166885
            Fax: +353 1 2837211
            Email: mary.e.logue@ucd.ie
            Class: plasmid ends.
FEATURES             Location/Qualifiers
     source          1. .752
                     /organism="Candida parapsilosis"
                     /mol_type="genomic DNA"
                     /strain="CLIB214"
                     /db_xref="taxon:5480"
                     /clone="cp19e05"
                     /clone_lib="Candida parapsilosis Random Genomic Library"
ORIGIN

Alignment Scores:
Pred. No.:                0.148          Length:         752
Score:                    84.00          Matches:        12
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       10             Gaps:           0

US-09-858-332G-15 (1-12) x CZ281894 (1-752)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        276 CACAACCACAACCACAACCACAACCACAACCACAAC 241

RESULT 100
BJ434227/c
LOCUS       BJ434227                759 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ434227 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv16c07 3', mRNA sequence.
ACCESSION   BJ434227
VERSION     BJ434227.1  GI:19408949
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 759)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .759
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv16c07"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:                0.15           Length:         759
Score:                    84.00          Matches:        12
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       3              Gaps:           0

US-09-858-332G-15 (1-12) x BJ434227 (1-759)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        208 CACAATCACAATCACAATCACAATCACAATCACAAT 173

RESULT 101
AG542244
LOCUS       AG542244                760 bp    DNA     linear   GSS 23-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-455M14.T7, genomic survey
            sequence.
ACCESSION   AG542244
VERSION     AG542244.1  GI:48302658
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 760)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : T7
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .760
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
```

```
                    /sub_species="molossinus"
                    /db_xref="taxon:57486"
                    /clone="MSMg01-455M14.T7"
                    /sex="male"
                    /tissue_type="mixture of kidney and spleen"
                    /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

Alignment Scores:
Pred. No.:              0.15          Length:        760
Score:                  84.00         Matches:       12
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     10            Gaps:          0

US-09-858-332G-15 (1-12) x AG542244 (1-760)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        123 CACAACCACAACCACAACCACAACCACAACCACAAC 158

RESULT 102
CO427845
LOCUS       CO427845                761 bp    mRNA    linear   EST 06-JUL-2004
DEFINITION  UI-M-HW0-crp-a-10-0-UI.r1 NIH_BMAP_HW0 Mus musculus cDNA clone
            IMAGE:30679209 5', mRNA sequence.
ACCESSION   CO427845
VERSION     CO427845.1  GI:49674139
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 761)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. James Lin University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Distribution information can be found at
            http://genome.uiowa.edu/distribution/mousefl.html
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES              Location/Qualifiers
     source           1. .761
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /strain="C57BL/6"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:30679209"
                      /tissue_type="whole eye"
                      /dev_stage="newborn (1, 5, 15 days )"
                      /lab_host="DH10B (T1 phage resistant)"
                      /clone_lib="NIH_BMAP_HW0"
                      /note="Organ: Eye; Vector: pYX-Asc; Site_1: EcoR I;
                      Site_2: Not I; The library was constructed according
                      Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                      1996. Denatured RNA was size fractionated on a 1% agarose
                      gel.First strand cDNA synthesis was primed with oligo-dT
                      primer containing a Not I site .Double strand cDNA was
                      size selected according to mRNA size fraction, ligated
                      with EcoR I adaptor, digested with NotI and then cloned
                      directionally into pYX-Asc vector. The library tag
                      sequence located between the Not I site and the polyA tail
                      is AATAATTACG. This library was created for the University
                      Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                      Developing Mouse Nervous System', supported by National
                      Institute of Mental Health (NIMH)."
ORIGIN

Alignment Scores:
Pred. No.:              0.15          Length:        761
Score:                  84.00         Matches:       12
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     7             Gaps:          0

US-09-858-332G-15 (1-12) x CO427845 (1-761)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        462 CACAACCACAACCACAACCACAACCACAACCACAAC 497

RESULT 103
BJ430651/c
LOCUS       BJ430651                762 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ430651 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv8k01 3', mRNA sequence.
ACCESSION   BJ430651
VERSION     BJ430651.1  GI:19405373
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 762)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES              Location/Qualifiers
     source           1. .762
                      /organism="Dictyostelium discoideum"
                      /mol_type="mRNA"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /clone="ddv8k01"
                      /sex="mat A"
                      /dev_stage="Growth phase"
                      /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:              0.15          Length:        762
Score:                  84.00         Matches:       12
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     3             Gaps:          0

US-09-858-332G-15 (1-12) x BJ430651 (1-762)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        203 CACAATCACAATCACAATCACAATCACAATCACAAT 168

RESULT 104
BJ431956/c
LOCUS       BJ431956                763 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ431956 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv17105 3', mRNA sequence.
ACCESSION   BJ431956
```

```
VERSION     BJ431956.1  GI:19406678
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 763)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .763
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv17l05"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:                0.151          Length:          763
Score:                    84.00          Matches:         12
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       3              Gaps:            0

US-09-858-332G-15 (1-12) x BJ431956 (1-763)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        201 CACAATCACAATCACAATCACAATCACAATCACAAT 166

RESULT 105
BJ432448/c
LOCUS       BJ432448                768 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ432448 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv18p15 3', mRNA sequence.
ACCESSION   BJ432448
VERSION     BJ432448.1  GI:19407170
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 768)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .768
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv18p15"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:                0.152          Length:          768
Score:                    84.00          Matches:         12
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       3              Gaps:            0

US-09-858-332G-15 (1-12) x BJ432448 (1-768)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        202 CACAATCACAATCACAATCACAATCACAATCACAAT 167

RESULT 106
BJ431981/c
LOCUS       BJ431981                774 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ431981 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv17a10 3', mRNA sequence.
ACCESSION   BJ431981
VERSION     BJ431981.1  GI:19406703
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 774)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .774
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv17a10"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:                0.153          Length:          774
Score:                    84.00          Matches:         12
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       3              Gaps:            0

US-09-858-332G-15 (1-12) x BJ431981 (1-774)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        206 CACAATCACAATCACAATCACAATCACAATCACAAT 171

RESULT 107
CO426813
LOCUS       CO426813                774 bp    mRNA    linear   EST 06-JUL-2004
DEFINITION  UI-M-HU0-crh-d-07-0-UI.r1 NIH_BMAP_HU0 Mus musculus cDNA clone
            IMAGE:30673902 5', mRNA sequence.
```

```
ACCESSION   CO426813
VERSION     CO426813.1  GI:49673107
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 774)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. James Lin University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Distribution information can be found at
            http://genome.uiowa.edu/distribution/mousefl.html
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1. .774
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:30673902"
                     /tissue_type="whole eye"
                     /dev_stage="newborn( 1, 5, 15 days )"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_HU0"
                     /note="Organ: Eye; Vector: pYX-Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel.First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site .Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is AATAATTACG. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH)."
ORIGIN

Alignment Scores:
Pred. No.:               0.153          Length:         774
Score:                   84.00          Matches:        12
Percent Similarity:      100.00%        Conservative:   0
Best Local Similarity:   100.00%        Mismatches:     0
Query Match:             100.00%        Indels:         0
DB:                      7              Gaps:           0

US-09-858-332G-15 (1-12) x CO426813 (1-774)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        479 CACAACCACAACCACAACCACAACCACAACCACAAC 514

RESULT 108
CO393602
LOCUS       CO393602                775 bp    mRNA    linear   EST 01-JUL-2004
DEFINITION  AGENCOURT_27552539 NIH_MGC_254 Rattus norvegicus cDNA clone
            IMAGE:7314222 5', mRNA sequence.
ACCESSION   CO393602
VERSION     CO393602.1  GI:49575518
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 775)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Drs. Josef Lazar & Howard Jacob, Medical
            College of Wisconsin
             cDNA Library Preparation: Express Genomics
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM15366  row: i  column: 04
            High quality sequence stop: 733.
FEATURES             Location/Qualifiers
     source          1. .775
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
                     /clone="IMAGE:7314222"
                     /sex="both"
                     /tissue_type="Brain - Pooled from several tissues from one
                     or more individuals"
                     /lab_host="DH10B TonA"
                     /clone_lib="NIH_MGC_254"
                     /note="Organ: brain/CNS; Vector: pExpress-1; Site_1:
                     EcoRV; Site_2: NotI; RNA obtained from brain tissue of 8
                     wk old animal. Tissues were snap-frozen and kept at -80C
                     before RNA extraction and purification (Tri-reagent
                     method). cDNA was primed using oligo-dT primer:
                     5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC(T)25-3' and cloned into
                     the EcoRV/NotI sites of pExpress-1. Size-selection >1.25kb
                     resulted in an average insert size of 2.18 kb. This
                     primary library is not normalized (normalized library is
                     NIH_MGC_255) and was constructed by Express Genomics
                     (Frederick, MD). Note: this is a NIH_MGC library"
ORIGIN

Alignment Scores:
Pred. No.:               0.153          Length:         775
Score:                   84.00          Matches:        12
Percent Similarity:      100.00%        Conservative:   0
Best Local Similarity:   100.00%        Mismatches:     0
Query Match:             100.00%        Indels:         0
DB:                      7              Gaps:           0

US-09-858-332G-15 (1-12) x CO393602 (1-775)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        240 CACAACCATAACCACAACCACAACCACAACCACAAC 275

RESULT 109
AG359323/c
LOCUS       AG359323                778 bp    DNA     linear   GSS 21-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-159F22.TJ, genomic survey
            sequence.
ACCESSION   AG359323
VERSION     AG359323.1  GI:47932633
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
```

```
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 778)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .778
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-159F22.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

Alignment Scores:
Pred. No.:                0.154          Length:          778
Score:                    84.00          Matches:         12
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       10             Gaps:            0

US-09-858-332G-15 (1-12) x AG359323 (1-778)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        265 CATAATCATAATCATAATCATAATCATAATCATAAT 230

RESULT 110
CK480095
LOCUS       CK480095                  783 bp    mRNA    linear   EST 14-JAN-2004
DEFINITION  AGENCOURT_17637714 NIH_MGC_235 Rattus norvegicus cDNA clone
            IMAGE:7106677 5', mRNA sequence.
ACCESSION   CK480095
VERSION     CK480095.1  GI:40824443
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 783)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Howard Jacobs
             cDNA Library Preparation: Express Genomics
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14967  row: m  column: 11
            High quality sequence stop: 619.
FEATURES             Location/Qualifiers
     source          1. .783
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
                     /clone="IMAGE:7106677"
                     /tissue_type="kidney, pooled"
                     /lab_host="DH10B TonA"
                     /clone_lib="NIH_MGC_235"
                     /note="Organ: kidney; Vector: pExpress-1; Site_1: EcoRV;
                     Site_2: NotI; RNA obtained from pooled kidney tissue from
                     a mix of male and female animals at 8 wk old. Tissues were
                     snap-frozen before RNA extraction and purification
                     (Tri-reagent method). cDNA was primed using oligo-dT
                     primer: 5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC(T)25-3' and
                     cloned into the EcoRV/NotI sites of pExpress-1.
                     Size-selection >1.4kb resulted in an average insert size
                     of 2.2 kb. This primary library is non-normalized
                     (normalized primary library is NIH_MGC_236) and was
                     constructed by Express Genomics (Frederick, MD). Note:
                     this is a NIH_MGC library."
ORIGIN

Alignment Scores:
Pred. No.:                0.155          Length:          783
Score:                    84.00          Matches:         12
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       7              Gaps:            0

US-09-858-332G-15 (1-12) x CK480095 (1-783)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        335 CACAACCACAACCACAACCACAACCACAACCACAAC 370

RESULT 111
CK596357
LOCUS       CK596357                  797 bp    mRNA    linear   EST 22-JAN-2004
DEFINITION  AGENCOURT_17837545 NIH_MGC_238 Rattus norvegicus cDNA clone
            IMAGE:7131493 5', mRNA sequence.
ACCESSION   CK596357
VERSION     CK596357.1  GI:41109440
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 797)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
```

Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Howard Jacobs
cDNA Library Preparation: Express Genomics
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLAM15032 row: g column: 11
High quality sequence stop: 654.

FEATURES Location/Qualifiers
source 1. .797
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="IMAGE:7131493"
/tissue_type="testis, pooled"
/lab_host="DH10B TonA"
/clone_lib="NIH_MGC_238"
/note="Organ: testis; Vector: pExpress-1; Site_1: EcoRV; Site_2: NotI; RNA obtained from testis tissue of 8 wk old animal. Tissues were snap-frozen and kept at -80C before RNA extraction and purification (Tri-reagent method). cDNA was primed using oligo-dT primer: 5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC(T)25-3' and cloned into the EcoRV/NotI sites of pExpress-1. Size-selection >1.4kb resulted in an average insert size of 1.9 kb. This primary library is normalized (non-normalized primary library is NIH_MGC_237) and was constructed by Express Genomics (Frederick, MD)"

ORIGIN

Alignment Scores:

Pred. No.:	0.158	Length:	797
Score:	84.00	Matches:	12
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	7	Gaps:	0

US-09-858-332G-15 (1-12) x CK596357 (1-797)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        307 CACAACCACAACCACAACCACAACCACAACCACAAC 342
```

RESULT 112
CO560704
LOCUS CO560704 802 bp mRNA linear EST 19-JUL-2004
DEFINITION AGENCOURT_28456958 NIH_MGC_247 Rattus norvegicus cDNA clone IMAGE:7323968 5', mRNA sequence.
ACCESSION CO560704
VERSION CO560704.1 GI:50373300
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 802)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Drs. Josef Lazar & Howard Jacob, Medical College of Wisconsin
cDNA Library Preparation: Open Biosystems
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLAM15391 row: o column: 06
High quality sequence start: 64
High quality sequence stop: 544.

FEATURES Location/Qualifiers
source 1. .802
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="IMAGE:7323968"
/lab_host="DH10B TonA"
/clone_lib="NIH_MGC_247"
/note="Organ: liver; Vector: pExpress-1; Site_1: EcoRV; Site_2: NotI; RNA obtained from testis tissue of 8 wk old animal. Tissues were snap-frozen and kept at -80C before RNA extraction and purification (Tri-reagent method). cDNA was primed using oligo-dT primer: 5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC(T)25-3' and cloned into the EcoRV/NotI sites of pExpress-1. Size-selection >1.25kb resulted in an average insert size of 1.6 kb. This primary library is normalized (non-normalized primary library is NIH_MGC_246) and was constructed by Open Biosystems. Note: this is a NIH_MGC library"

ORIGIN

Alignment Scores:

Pred. No.:	0.159	Length:	802
Score:	84.00	Matches:	12
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	7	Gaps:	0

US-09-858-332G-15 (1-12) x CO560704 (1-802)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        224 CACAACCACAACCACAACCACAACCACAACCACAAC 259
```

RESULT 113
CB313044
LOCUS CB313044 811 bp mRNA linear EST 04-MAR-2003
DEFINITION AGENCOURT_11467474 NICHD_Rr_Pit1 Rattus norvegicus cDNA clone IMAGE:6890909 5', mRNA sequence.
ACCESSION CB313044
VERSION CB313044.1 GI:28835934
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 811)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: John C. Marshall, M.D., Ph.D
cDNA Library Preparation: CLONTECH
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLCM3151 row: c column: 04
High quality sequence stop: 307.

FEATURES Location/Qualifiers
source 1. .811
/organism="Rattus norvegicus"

/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="IMAGE:6890909"
/tissue_type="Pitutary"
/lab_host="DH10B"
/clone_lib="NICHD_Rr_Pit1"
/note="Vector: pDNR-LIB; Site_1: SfiI; Site_2: SfiI; 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.23 kb (range 0.5-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

ORIGIN

Alignment Scores:

Pred. No.:	0.161	Length:	811
Score:	84.00	Matches:	12
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-09-858-332G-15 (1-12) x CB313044 (1-811)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        303 CACAACCATAACCACAACCACAACCACAACCACAAC 338
```

RESULT 114
DN475931
LOCUS DN475931 811 bp mRNA linear EST 09-MAR-2005
DEFINITION altr202xg17 A. brassicicola mycelial culture grown under conditions of Nitrogen Starvation Alternaria brassicicola cDNA clone altr202xg17, mRNA sequence.
ACCESSION DN475931
VERSION DN475931.1 GI:60673242
KEYWORDS EST.
SOURCE Alternaria brassicicola
ORGANISM Alternaria brassicicola
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes; Pleosporales; Pleosporaceae; mitosporic Pleosporaceae; Alternaria.
REFERENCE 1 (bases 1 to 811)
AUTHORS Cramer,R.A., Craven,K.D., Thon,M.R., Cho,Y., Knudson,D.L., Mitchell,T.K. and Lawrence,C.B.
TITLE Expressed Sequence Tag (EST) Analysis of a Compatible Alternaria brassicicola-Brassica oleracea Interaction
JOURNAL Unpublished (2005)
COMMENT Contact: Dr. Thomas K. Mitchell
Center for Integrated Fungal Research, NC State University
851 Main Campus Dr. Suite 233, Raleigh, NC 27606, USA
Tel: (919) 513-3926
Fax: (919) 513-0024
Email: thomas_mitchell@ncsu.edu
Seq primer: T7 SP6 primer.
FEATURES Location/Qualifiers
source 1. .811
/organism="Alternaria brassicicola"
/mol_type="mRNA"
/strain="ATCC 96866"
/db_xref="taxon:29001"
/clone="altr202xg17"
/dev_stage="mycelia"
/clone_lib="A. brassicicola mycelial culture grown under conditions of Nitrogen Starvation"
/note="Vector: pGEMt; Mycelial culture grown under conditions of Nitrogen Starvation."

ORIGIN

Alignment Scores:

Pred. No.:	0.161	Length:	811
Score:	84.00	Matches:	12
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	8	Gaps:	0

US-09-858-332G-15 (1-12) x DN475931 (1-811)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        163 CACAACCACAACCACAACCACAACCACAACCACAAC 198
```

RESULT 115
CK469853
LOCUS CK469853 832 bp mRNA linear EST 13-JAN-2004
DEFINITION AGENCOURT_17644694 NIH_MGC_237 Rattus norvegicus cDNA clone IMAGE:7113432 5', mRNA sequence.
ACCESSION CK469853
VERSION CK469853.1 GI:40808335
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 832)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Howard Jacobs
cDNA Library Preparation: Express Genomics
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM14985 row: f column: 22
High quality sequence start: 175
High quality sequence stop: 675.
FEATURES Location/Qualifiers
source 1. .832
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="IMAGE:7113432"
/tissue_type="testis, pooled"
/lab_host="DH10B TonA"
/clone_lib="NIH_MGC_237"
/note="Organ: testis; Vector: pExpress-1; Site_1: EcoRV; Site_2: NotI; RNA obtained from testis tissue of 8 wk old animal. Tissues were snap-frozen and kept at -80C before RNA extraction and purification (Tri-reagent method). cDNA was primed using oligo-dT primer: 5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC(T)25-3' and cloned into the EcoRV/NotI sites of pExpress-1. Size-selection >1.4kb resulted in an average insert size of 2.4 kb. This primary library is not normalized (normalized primary library is NIH_MGC_238) and was constructed by Express Genomics (Frederick, MD)"

ORIGIN

Alignment Scores:

Pred. No.:	0.165	Length:	832
Score:	84.00	Matches:	12
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	7	Gaps:	0

```
US-09-858-332G-15 (1-12) x CK469853 (1-832)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db         42 CACAACCACAACCACAACCACAACCACAACCACAAC 77

RESULT 116
CO396373
LOCUS       CO396373                 836 bp    mRNA    linear   EST 01-JUL-2004
DEFINITION  AGENCOURT_27851354 NIH_MGC_252 Rattus norvegicus cDNA clone
            IMAGE:7310835 5', mRNA sequence.
ACCESSION   CO396373
VERSION     CO396373.1  GI:49578289
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 836)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Drs. Josef Lazar & Howard Jacob, Medical
            College of Wisconsin
             cDNA Library Preparation: Express Genomics
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM15357  row: l  column: 01
            High quality sequence start: 8
            High quality sequence stop: 674.
FEATURES             Location/Qualifiers
     source          1. .836
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
                     /clone="IMAGE:7310835"
                     /tissue_type="Ovary, strain - Norway Line3. Age8 weeks.
                     Tissue was snap-frozen adn transferred in -70C. RNAse free
                     for the entire procedure"
                     /lab_host="DH10B TonA"
                     /clone_lib="NIH_MGC_252"
                     /note="Organ: ovary; Vector: pExpress1; Site_1: NotI;
                     Site_2: NotI; RNA obtained from female overies animals at
                     8 wk old. Tissues were snap-frozen and kept at -80C for
                     two days before RNA extraction and purification
                     (Tri-reagent method). cDNA was primed using oligo-dT
                     primer: 5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC(T)25-3' and
                     cloned into the EcoRV/NotI sites of pExpress-1.
                     Size-selection >1.25kb resulted in an average insert size
                     of 1.7kb. This primary library is not normalized
                     (normalized library is NIH_MGC_252) and was constructed by
                     Express Genomics (Frederick, MD). Note: this is a NIH_MGC
                     library"
ORIGIN

Alignment Scores:
Pred. No.:              0.166          Length:          836
Score:                  84.00          Matches:         12
Percent Similarity:     100.00%        Conservative:    0
Best Local Similarity:  100.00%        Mismatches:      0
Query Match:            100.00%        Indels:          0
DB:                     7              Gaps:            0
```

```
US-09-858-332G-15 (1-12) x CO396373 (1-836)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        210 CACAACCACAACCACAACCACAACCACAACCACAAC 245

RESULT 117
CK470094
LOCUS       CK470094                 854 bp    mRNA    linear   EST 13-JAN-2004
DEFINITION  AGENCOURT_17647746 NIH_MGC_237 Rattus norvegicus cDNA clone
            IMAGE:7113899 5', mRNA sequence.
ACCESSION   CK470094
VERSION     CK470094.1  GI:40808576
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 854)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Howard Jacobs
             cDNA Library Preparation: Express Genomics
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14986  row: j  column: 09
            High quality sequence stop: 677.
FEATURES             Location/Qualifiers
     source          1. .854
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
                     /clone="IMAGE:7113899"
                     /tissue_type="testis, pooled"
                     /lab_host="DH10B TonA"
                     /clone_lib="NIH_MGC_237"
                     /note="Organ: testis; Vector: pExpress-1; Site_1: EcoRV;
                     Site_2: NotI; RNA obtained from testis tissue of 8 wk old
                     animal. Tissues were snap-frozen and kept at -80C before
                     RNA extraction and purification (Tri-reagent method). cDNA
                     was primed using oligo-dT primer:
                     5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC(T)25-3' and cloned into
                     the EcoRV/NotI sites of pExpress-1. Size-selection >1.4kb
                     resulted in an average insert size of 2.4 kb. This primary
                     library is not normalized (normalized primary library is
                     NIH_MGC_238) and was constructed by Express Genomics
                     (Frederick, MD)"
ORIGIN

Alignment Scores:
Pred. No.:              0.169          Length:          854
Score:                  84.00          Matches:         12
Percent Similarity:     100.00%        Conservative:    0
Best Local Similarity:  100.00%        Mismatches:      0
Query Match:            100.00%        Indels:          0
DB:                     7              Gaps:            0

US-09-858-332G-15 (1-12) x CK470094 (1-854)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        502 CACAACCACAACCACAACCACAACCACAACCACAAC 537
```

```
RESULT 118
BF384538
LOCUS       BF384538                898 bp    mRNA    linear   EST 27-NOV-2000
DEFINITION  602046807F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:4196520 5',
            mRNA sequence.
ACCESSION   BF384538
VERSION     BF384538.1  GI:11365843
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 898)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9533  row: m  column: 01
            High quality sequence stop: 727.
FEATURES             Location/Qualifiers
     source          1. .898
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4196520"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Li9"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.9 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
ORIGIN

Alignment Scores:
Pred. No.:                0.179          Length:         898
Score:                    84.00          Matches:        12
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       2              Gaps:           0

US-09-858-332G-15 (1-12) x BF384538 (1-898)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        185 CACAACCACAACCACAACCACAACCACAACCACAAC 220

RESULT 119
BI247099
LOCUS       BI247099                913 bp    mRNA    linear   EST 17-JUL-2001
DEFINITION  602960664F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:5126270 5',
            mRNA sequence.
ACCESSION   BI247099
VERSION     BI247099.1  GI:14791697
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 913)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11308  row: p  column: 15
            High quality sequence stop: 830.
FEATURES             Location/Qualifiers
     source          1. .913
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5126270"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Li9"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.9 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
ORIGIN

Alignment Scores:
Pred. No.:                0.182          Length:         913
Score:                    84.00          Matches:        12
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       2              Gaps:           0

US-09-858-332G-15 (1-12) x BI247099 (1-913)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db         21 CACAACCACAACCACAACCACAACCACAACCACAAC 56

RESULT 120
BI147575
LOCUS       BI147575                924 bp    mRNA    linear   EST 05-JUL-2001
DEFINITION  602913017F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:5053732 5',
            mRNA sequence.
ACCESSION   BI147575
VERSION     BI147575.1  GI:14607576
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 924)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11146  row: b  column: 05
            High quality sequence stop: 601.
FEATURES             Location/Qualifiers
     source          1. .924
                     /organism="Mus musculus"
                     /mol_type="mRNA"
```

```
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5053732"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Li9"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.9 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
ORIGIN

Alignment Scores:
Pred. No.:              0.184          Length:        924
Score:                  84.00          Matches:       12
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     2              Gaps:          0

US-09-858-332G-15 (1-12) x BI147575 (1-924)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db         38 CACAACCACAACCACAACCACAACCACAACCACAAC 73

RESULT 121
BQ946631
LOCUS       BQ946631                929 bp    mRNA    linear   EST 21-AUG-2002
DEFINITION  AGENCOURT_8931119 NIH_MGC_94 Mus musculus cDNA clone IMAGE:6468248
            5', mRNA sequence.
ACCESSION   BQ946631
VERSION     BQ946631.1  GI:22362109
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 929)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13995  row: h  column: 09
            High quality sequence stop: 622.
FEATURES             Location/Qualifiers
     source          1. .929
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6468248"
                     /tissue_type="retina"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_94"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 3.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
ORIGIN

Alignment Scores:
Pred. No.:              0.185          Length:        929
Score:                  84.00          Matches:       12
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     5              Gaps:          0

US-09-858-332G-15 (1-12) x BQ946631 (1-929)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        137 CACAACCACAACCACAACCACAACCACAACCACAAC 172

RESULT 122
AG403889/c
LOCUS       AG403889                938 bp    DNA     linear   GSS 21-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-261I10.TJ, genomic survey
            sequence.
ACCESSION   AG403889
VERSION     AG403889.1  GI:48046575
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 938)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .938
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-261I10.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

Alignment Scores:
Pred. No.:              0.187          Length:        938
Score:                  84.00          Matches:       12
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     10             Gaps:          0
```

```
US-09-858-332G-15 (1-12) x AG403889 (1-938)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        242 CATAATCATAATCATAATCATAATCATAATCATAAT 207

RESULT 123
BI145320
LOCUS       BI145320                941 bp    mRNA    linear   EST 05-JUL-2001
DEFINITION  602910458F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:5051463 5',
            mRNA sequence.
ACCESSION   BI145320
VERSION     BI145320.1  GI:14605321
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 941)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11140  row: c  column: 16
            High quality sequence stop: 861.
FEATURES             Location/Qualifiers
     source          1. .941
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5051463"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Li9"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.9 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
ORIGIN

Alignment Scores:
Pred. No.:                0.188          Length:          941
Score:                    84.00          Matches:         12
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       2              Gaps:            0

US-09-858-332G-15 (1-12) x BI145320 (1-941)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db         48 CACAACCACAACCACAACCACAACCACAACCACAAC 83

RESULT 124
BF384348
LOCUS       BF384348                963 bp    mRNA    linear   EST 27-NOV-2000
DEFINITION  602046701F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:4195872 5',
            mRNA sequence.
ACCESSION   BF384348
VERSION     BF384348.1  GI:11365653
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 963)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9532  row: b  column: 01
            High quality sequence stop: 719.
FEATURES             Location/Qualifiers
     source          1. .963
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4195872"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Li9"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.9 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
ORIGIN

Alignment Scores:
Pred. No.:                0.192          Length:          963
Score:                    84.00          Matches:         12
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       2              Gaps:            0

US-09-858-332G-15 (1-12) x BF384348 (1-963)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        157 CACAACCACAACCACAACCACAACCACAACCACAAC 192

RESULT 125
CNS01S89
LOCUS       CNS01S89                969 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            169I16 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL164898
VERSION     AL164898.1  GI:7802636
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  PUBMED    10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
```

```
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
   PUBMED   10899143
REFERENCE   3  (bases 1 to 969)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .969
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="169I16"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG169BE08LP1
                     end : T7"
ORIGIN

Alignment Scores:
Pred. No.:              0.193          Length:         969
Score:                  84.00          Matches:        12
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     10             Gaps:           0

US-09-858-332G-15 (1-12) x CNS01S89 (1-969)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        220 CATAATCATAATCATAATCATAATCATAATCATAAT 255

RESULT 126
CNS0A3V5/c
LOCUS       CNS0A3V5                983 bp    mRNA    linear   HTC 06-FEB-2004
DEFINITION  Arabidopsis thaliana Full-length cDNA Complete sequence from clone
            GSLTPGH90ZD08 of Hormone Treated Callus of strain col-0 of
            Arabidopsis thaliana (thale cress).
ACCESSION   BX828453
VERSION     BX828453.1  GI:42461786
KEYWORDS    HTC; GSLT_cDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 983)
  AUTHORS   Castelli,V., Aury,J.M., Jaillon,O., Wincker,P., Clepet,C.,
            Menard,M., Cruaud,C., Quetier,F., Scarpelli,C., Schachter,V.,
            Temple,G., Caboche,M., Weissenbach,J. and Salanoubat,M.
  TITLE     Whole Genome Sequence Comparisons and 'Full-Length' cDNA Sequences:
            A Combined Approach to Evaluate and Improve Arabidopsis Genome
            Annotation
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 983)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-2003) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     The sequences are based on single pass reads.
            Life Technologies (a division of Invitrogen) members carried out
            full-length librairies construction : Temple G.
            Genoscope members carried out sequencing and annotation : Castelli
            V., Aury J.M., Jaillon O., Wincker P., Menard M., Cruaud C.,
            Schachter V., Weissenbach J., Salanoubat M.
            URGV INRA : Clepet C., Caboche M.
            Annotation is based on the June 2003 version of the Arabidopsis
            genome released by MIPS (Munich Information center for Protein
            Sequences). 5 prime and 3 prime are assembled with Phrap.
            http://www.genoscope.cns.fr/externe/sequences/Banque_Projet_EF/Full
            _length
            http://www.genoscope.cns.fr/cgi-bin/ggb/ggb?source=Arabidopsis.
FEATURES             Location/Qualifiers
     source          1. .983
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="GSLTPGH90ZD08"
                     /tissue_type="Hormone Treated Callus"
                     /ecotype="Col-0"
                     /plasmid="pCMVSPORT_6"
     gene            complement(1. .983)
                     /gene="At4g21810"
ORIGIN

Alignment Scores:
Pred. No.:              0.196          Length:         983
Score:                  84.00          Matches:        12
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     4              Gaps:           0

US-09-858-332G-15 (1-12) x CNS0A3V5 (1-983)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        108 CACAATCACAATCACAATCACAATCACAATCACAAT 73

RESULT 127
CW926142
LOCUS       CW926142                1127 bp    DNA     linear   GSS 24-FEB-2005
DEFINITION  EDCBE62TF A. castellanii, 6-8 kb library from total genomic DNA
            Acanthamoeba castellanii genomic clone EDCBE62, genomic survey
            sequence.
ACCESSION   CW926142
VERSION     CW926142.1  GI:60236012
KEYWORDS    GSS.
SOURCE      Acanthamoeba castellanii
  ORGANISM  Acanthamoeba castellanii
            Eukaryota; Acanthamoebidae; Acanthamoeba.
REFERENCE   1  (bases 1 to 1127)
  AUTHORS   Anderson,I.J. and Loftus,B.J.
  TITLE     Gene discovery in the Acanthamoeba castellanii genome
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Iain Anderson
            The Institute for Genomic Research (TIGR; www.tigr.org)
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-795-7949
            Fax: 301-838-0208
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .1127
                     /organism="Acanthamoeba castellanii"
                     /mol_type="genomic DNA"
                     /strain="Neff"
                     /db_xref="taxon:5755"
                     /clone="EDCBE62"
                     /clone_lib="A. castellanii, 6-8 kb library from total
                     genomic DNA"
                     /note="Vector: pHOS2"
ORIGIN

Alignment Scores:
Pred. No.:              0.226          Length:         1127
Score:                  84.00          Matches:        12
Percent Similarity:     100.00%        Conservative:   0
```

```
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     10             Gaps:          0

US-09-858-332G-15 (1-12) x CW926142 (1-1127)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        776 CACAACCACAACCACAACCACAACCACAACCACAAC 811

RESULT 128
CG748246/c
LOCUS       CG748246                1147 bp    DNA     linear   GSS 24-OCT-2003
DEFINITION  P042-1-D01.za Ppa EcoRI BAC Library Pristionchus pacificus genomic,
            genomic survey sequence.
ACCESSION   CG748246
VERSION     CG748246.1  GI:37969172
KEYWORDS    GSS.
SOURCE      Pristionchus pacificus
  ORGANISM  Pristionchus pacificus
            Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
            Neodiplogasteridae; Pristionchus.
REFERENCE   1  (bases 1 to 1147)
  AUTHORS   Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
            Buntjer,J., van der Meulen,M. and Sommer,R.J.
  TITLE     An integrated physical and genetic map of the nematode Pristionchus
            pacificus
  JOURNAL   Mol. Genet. Genomics 269 (5), 715-722 (2003)
   PUBMED   12884007
COMMENT     Contact: Sommer RJ
            Evolutionary Biology
            Max-Planck-Institute for Developmental Biology
            Spemannstr. 37-39, Tuebingen D-72076, Germany
            Tel: 00497071601371
            Fax: 00497071601498
            Email: ralf.sommer@tuebingen.mpg.de
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .1147
                     /organism="Pristionchus pacificus"
                     /mol_type="genomic DNA"
                     /strain="California"
                     /db_xref="taxon:54126"
                     /clone_lib="Ppa EcoRI BAC Library"
                     /note="The library was generated by a partial digest of
                     the genomic DNA with EcoRI and cloning into the BAC
                     vector."
ORIGIN

Alignment Scores:
Pred. No.:              0.231          Length:        1147
Score:                  84.00          Matches:       12
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     10             Gaps:          0

US-09-858-332G-15 (1-12) x CG748246 (1-1147)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        334 CATAATCATAATCATAATCATAATCATAATCATAAT 299

RESULT 129
CG751785
LOCUS       CG751785                1260 bp    DNA     linear   GSS 24-OCT-2003
DEFINITION  P046-2-C07.ya Ppa EcoRI BAC Library Pristionchus pacificus genomic,
            genomic survey sequence.
ACCESSION   CG751785
VERSION     CG751785.1  GI:37974592
KEYWORDS    GSS.
SOURCE      Pristionchus pacificus
  ORGANISM  Pristionchus pacificus
            Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
            Neodiplogasteridae; Pristionchus.
REFERENCE   1  (bases 1 to 1260)
  AUTHORS   Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
            Buntjer,J., van der Meulen,M. and Sommer,R.J.
  TITLE     An integrated physical and genetic map of the nematode Pristionchus
            pacificus
  JOURNAL   Mol. Genet. Genomics 269 (5), 715-722 (2003)
   PUBMED   12884007
COMMENT     Contact: Sommer RJ
            Evolutionary Biology
            Max-Planck-Institute for Developmental Biology
            Spemannstr. 37-39, Tuebingen D-72076, Germany
            Tel: 00497071601371
            Fax: 00497071601498
            Email: ralf.sommer@tuebingen.mpg.de
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .1260
                     /organism="Pristionchus pacificus"
                     /mol_type="genomic DNA"
                     /strain="California"
                     /db_xref="taxon:54126"
                     /clone_lib="Ppa EcoRI BAC Library"
                     /note="The library was generated by a partial digest of
                     the genomic DNA with EcoRI and cloning into the BAC
                     vector."
ORIGIN

Alignment Scores:
Pred. No.:              0.255          Length:        1260
Score:                  84.00          Matches:       12
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     10             Gaps:          0

US-09-858-332G-15 (1-12) x CG751785 (1-1260)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        528 CATAATCATAATCATAATCATAATCATAATCATAAT 563

RESULT 130
CO484355
LOCUS       CO484355                 228 bp    mRNA    linear   EST 29-JUN-2005
DEFINITION  GQ0204.B3_E17 GQ020: Clean ROOTS systems -Diurnal cycle Picea
            glauca cDNA clone GenomeQuebec_Id:GQ0204E17 5', mRNA sequence.
ACCESSION   CO484355
VERSION     CO484355.1  GI:50167736
KEYWORDS    EST.
SOURCE      Picea glauca (white spruce)
  ORGANISM  Picea glauca
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
REFERENCE   1  (bases 1 to 228)
  AUTHORS   Morency,M.-J., Cooke,J., Pavy,N., Parsons,L., Paule,C., Seguin,A.,
            Retzel,E., Butterfield,Y., Barber,S., Yang,G., Stott,J.,
            Siddiqui,A., Holt,R., Marra,M. and MacKay,J.
  TITLE     Arborea EST sequencing in Picea glauca (white spruce)
  JOURNAL   Unpublished (2004)
COMMENT     Contact: John MacKay
            Centre de Recherche en Biologie Forestiere
            Universite Laval
            Pavillon Charles-Eugene Marchand, Quebec, Quebec, CANADA G1K 7P4
            Fax: 418 656 7493
            Email: jmackay@rsvs.ulaval.ca
            Center for Computational Genomics and Bioinformatics (CCGB),
            University of Minnesota, MN_id Identifier: MN5235471 Clone ID:
            GQ0204_E17 Clones available through: John MacKay, Ph. D. Professeur
            adjoint -Assistant professor EMAIL: jmackay@rsvs.ulaval.ca Centre
            de Recherche en Biologie Forestiere (Forest Biology Research
            Center) Universite Laval Quebec, Quebec CANADA G1K 7P4
```

```
          Plate: 4  row: 17  column: E
          Seq primer: T3 Primer.
FEATURES             Location/Qualifiers
     source          1. .228
                     /organism="Picea glauca"
                     /mol_type="mRNA"
                     /strain="pg-653"
                     /db_xref="taxon:3330"
                     /clone="GenomeQuebec_Id:GQ0204E17"
                     /sex="Hermaphrodite"
                     /tissue_type="Whole roots systems, predominantly free of
                     michorizee"
                     /lab_host="E. coli DH10B cells"
                     /clone_lib="GQ020: Clean ROOTS systems -Diurnal cycle"
                     /note="Organ: whole root systems form 15 cm tall, 4 month
                     old seedlings produced by somatic embryogenesis; Vector:
                     pBluescript II SK (+) XR; Site_1: Eco-RI; Site_2: Xho-I;
                     Tissues were pooled from several plants grown under 20h
                     light/4h dark and collected every 2 hours over a period of
                     28 hours (4 dark time points and 11 time points in light).
                     cDNA was prepared from 5 mg of poly A+ selected RNA and
                     was directionally ligated into the pBluescript II SK (+)
                     XR vector (Stratagene), transformed by electroporation
                     into DH10B cells (In vitrogen) for propagation"
ORIGIN

Alignment Scores:
Pred. No.:              0.188          Length:         228
Score:                  79.00          Matches:        11
Percent Similarity:     100.00%        Conservative:   1
Best Local Similarity:  91.67%         Mismatches:     0
Query Match:            94.05%         Indels:         0
DB:                     7              Gaps:           0

US-09-858-332G-15 (1-12) x CO484355 (1-228)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||:::
Db         88 CACAACCACAACCACAACCACAACCACAACCACAGC 123

RESULT 131
CD852815/c
LOCUS       CD852815                357 bp    mRNA    linear   EST 11-JUL-2003
DEFINITION  DH0ALL9ZH02ZZM1 HaDevR5 Helianthus annuus cDNA clone HaDevR59H02,
            mRNA sequence.
ACCESSION   CD852815
VERSION     CD852815.1  GI:32536631
KEYWORDS    EST.
SOURCE      Helianthus annuus (common sunflower)
  ORGANISM  Helianthus annuus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; campanulids; Asterales; Asteraceae; Asteroideae;
            Heliantheae; Helianthus.
REFERENCE   1  (bases 1 to 357)
  AUTHORS   Genoplante.
  TITLE     Genoplante, a major partnership french program in plant genomics
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Genoplante
            Genoplante
            93, rue Henri Rochefort 91025 EVRY CEDEX France
            Tel: 33 1 69 47 54 00
            Fax: 33 1 69 47 54 10
            This sequence has been generated in the framework of the french
            plant genomics programme 'Genoplante' (http://www.genoplante.com
            and http://genoplante-info.infobiogen.fr).
FEATURES             Location/Qualifiers
     source          1. .357
                     /organism="Helianthus annuus"
                     /mol_type="mRNA"
                     /cultivar="psc8"
                     /db_xref="taxon:4232"
                     /clone="HaDevR59H02"
                     /tissue_type="4 days after self-pollination embryo"
                     /clone_lib="HaDevR5"
ORIGIN

Alignment Scores:
Pred. No.:              0.3            Length:         357
Score:                  79.00          Matches:        11
Percent Similarity:     100.00%        Conservative:   1
Best Local Similarity:  91.67%         Mismatches:     0
Query Match:            94.05%         Indels:         0
DB:                     6              Gaps:           0

US-09-858-332G-15 (1-12) x CD852815 (1-357)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||:::
Db        342 CATAATCATAATCATAATCATAATCATAATCATAGT 307

RESULT 132
BM300074
LOCUS       BM300074                358 bp    mRNA    linear   EST 22-JAN-2002
DEFINITION  MCR054H05_24676 Ice plant Lambda Uni-Zap XR expression library, 48
            hours NaCl treatment prescreened for removal of highly abundant
            transcripts Mesembryanthemum crystallinum cDNA clone MCR054H05 5,
            mRNA sequence.
ACCESSION   BM300074
VERSION     BM300074.1  GI:18022449
KEYWORDS    EST.
SOURCE      Mesembryanthemum crystallinum (common iceplant)
  ORGANISM  Mesembryanthemum crystallinum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            Caryophyllales; Aizoaceae; Mesembryanthemum.
REFERENCE   1  (bases 1 to 358)
  AUTHORS   Cushman, J.C.
  TITLE     An expressed sequence tag database for the common ice plant,
            Mesembryanthemum crystallinum
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Cushman JC
            Department of Biochemistry
            University of Nevada
            MS200, Reno, NV 89557-0014, USA
            Tel: 775-784-1918
            Fax: 775-784-1650
            Email: jcushman@unr.edu
            PCR PRimers
            FORWARD: T3 20mer
            BACKWARD: T7 21mer
            Plate: 054  row: H  column: 05
            Seq primer: T3 20mer
            High quality sequence stop: 358.
FEATURES             Location/Qualifiers
     source          1. .358
                     /organism="Mesembryanthemum crystallinum"
                     /mol_type="mRNA"
                     /db_xref="taxon:3544"
                     /clone="MCR054H05"
                     /tissue_type="leaf"
                     /dev_stage="six-week-old"
                     /clone_lib="Ice plant Lambda Uni-Zap XR expression
                     library, 48 hours NaCl treatment prescreened for removal
                     of highly abundant transcripts"
                     /note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI"
ORIGIN

Alignment Scores:
Pred. No.:              0.301          Length:         358
Score:                  79.00          Matches:        11
Percent Similarity:     100.00%        Conservative:   1
Best Local Similarity:  91.67%         Mismatches:     0
Query Match:            94.05%         Indels:         0
DB:                     3              Gaps:           0
```

```
US-09-858-332G-15 (1-12) x BM300074 (1-358)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||||||||||||||||||||||||||||||
Db         10 CATCACCACAACCACAACCACAACCACAACCACAAC 45

RESULT 133
BH142412/c
LOCUS       BH142412                363 bp    DNA     linear   GSS 16-AUG-2001
DEFINITION  TDGDO50TH cTOG Lycopersicon esculentum genomic clone cTOG24I3,
            genomic survey sequence.
ACCESSION   BH142412
VERSION     BH142412.1  GI:15194980
KEYWORDS    GSS.
SOURCE      Lycopersicon esculentum (Solanum lycopersicum)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 363)
  AUTHORS   van der Hoeven,R., Sun,H., Cho,J., Utterback,T., Ronning,C. and
            Tanksley,S.
  TITLE     Tomato Demethylated Genomic DNA Sequences
  JOURNAL   Unpublished (2001)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html
            tomato demethylated genomic DNA
            Insert Length: 1270   Std Error: 0.00
            Seq primer: M13F-R
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .363
                     /organism="Lycopersicon esculentum"
                     /mol_type="genomic DNA"
                     /cultivar="E6203"
                     /db_xref="taxon:4081"
                     /clone="cTOG24I3"
                     /tissue_type="young leaves"
                     /dev_stage="12-14 weeks post harvest"
                     /lab_host="E.coli JM109"
                     /clone_lib="cTOG"
                     /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                     XhoI; This library was made from short EcoRI digested
                     fragments of the genome of Lycopersicon esculentum ligated
                     into pBS (SK-). The fragments were  cloned into the
                     methylation restrictive E.coli strain JM109 with the
                     purpose of enriching the library for non-methylated DNA
                     fragments.  This procedure may enrich the pool of cloned
                     fragments in JM109 cells for sequences representing
                     expressed genes. Average insert size 1.27 kb."
ORIGIN

Alignment Scores:
Pred. No.:                0.305          Length:          363
Score:                    79.00          Matches:         11
Percent Similarity:       100.00%        Conservative:    1
Best Local Similarity:    91.67%         Mismatches:      0
Query Match:              94.05%         Indels:          0
DB:                       9              Gaps:            0

US-09-858-332G-15 (1-12) x BH142412 (1-363)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||||||||||||:::|||||||||||||||||||
Db        232 CACAACCATAATCATAGTCATAATCATAATCATAAT 197

RESULT 134
BP805483/c
LOCUS       BP805483                427 bp    mRNA    linear   EST 15-FEB-2005
DEFINITION  BP805483 RAFL16 Arabidopsis thaliana cDNA clone RAFL24-06-C01 5',
            mRNA sequence.
ACCESSION   BP805483
VERSION     BP805483.1  GI:59841562
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 427)
  AUTHORS   Seki,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M., Akiyama,K.,
            Iida,K., Enju,A., Sakurai,T., Arakawa,T., Carninci,P., Fukuda,S.,
            Iida,J., Kawai,J., Sasaki,D., Shiraki,T., Hayashizaki,Y. and
            Shinozaki,K.
  TITLE     Large-Scale Analysis of RIKEN Arabidopsis Full-length cDNAs
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            An Arabidopsis full-length cDNA library was constructed essentially
            as reported previously (Seki et al., 1998, 2002).This clone is in a
            modified pBluescript vector.
            Please visit our web site (http://pfgweb.gsc.riken.jp and
            http://rarge.gsc.riken.jp) for further details.
FEATURES             Location/Qualifiers
     source          1. .427
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="RAFL24-06-C01"
                     /lab_host="DH10B"
                     /clone_lib="RAFL16"
                     /note="Site_1: BamHI; Site_2: SalI; dark-grown"
ORIGIN

Alignment Scores:
Pred. No.:                0.361          Length:          427
Score:                    79.00          Matches:         11
Percent Similarity:       100.00%        Conservative:    1
Best Local Similarity:    91.67%         Mismatches:      0
Query Match:              94.05%         Indels:          0
DB:                       3              Gaps:            0

US-09-858-332G-15 (1-12) x BP805483 (1-427)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||||||||||||||||||||||||||||||:::
Db        184 CACAATCACAATCACAATCACAATCACAATCACCAT 149

RESULT 135
CA890073/c
LOCUS       CA890073                510 bp    mRNA    linear   EST 20-DEC-2002
DEFINITION  B0156H06-5N NIA Mouse Neural Stem Cell (Differentiated) cDNA
            Library (Long) Mus musculus cDNA clone NIA:B0156H06 IMAGE:30099257
            5', mRNA sequence.
ACCESSION   CA890073
VERSION     CA890073.1  GI:27341622
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 510)
  AUTHORS   Piao,Y., Dudekula,D.B., Qian,Y., Martin,P.R., Aiba,K., Vescovi,A.L.
            and Ko,M.S.H.
  TITLE     Systematic Analyses of NIA Mouse Neural Stem Cell (Differentiated)
            cDNA Library (Long)
```

```
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Dawood B. Dudekula
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgsun.grc.nia.nih.gov
            Plate: B0156  row: H  column: 06
            Seq primer: -21M13 Reverse
            High quality sequence stop: 510
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .510
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="CD1"
                     /db_xref="niaEST:B0156H06-5N"
                     /db_xref="taxon:10090"
                     /clone="NIA:B0156H06 IMAGE:30099257"
                     /dev_stage="Adult"
                     /lab_host="DH10B"
                     /clone_lib="NIA Mouse Neural Stem Cell (Differentiated)
                     cDNA Library (Long)"
                     /note="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2:
                     NotI; Mouse cDNA project by the Laboratory of Genetics,
                     National Institute on Aging (NIA), Intramural Research
                     Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is
                     a long-transcript enriched cDNA library (Ref. Genome Res.
                     11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were
                     obtained from Dr. Angelo L. Vescovi (Institute for Stem
                     Cell Research, Italy). Double-stranded cDNAs were
                     synthesized with an Oligo(dT) primer [Invitrogen:
                     5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTTT-3'] from
                     2.0 Microgram of total RNA, treated with T4 DNA
                     polymerase, and purified by ethanol-precipitation. The
                     cDNAs were ligated to Lone-linker LL-Sal4, purified by
                     phenol/chloroform, and separated from free linkers by
                     Centricon 100. Then, the cDNAs were amplified by
                     long-range high fidelity PCR using Ex Taq polymerase
                     (Takara) with a primer Sal4-S. The products were purified
                     by phenol/chloroform and Centricon 100. The cDNAs were
                     digested with SalI and NotI enzymes and cloned into
                     SalI/NotI site of pSPORT1 plasmid vector. The DH10B E.
                     coli host was transformed with the ligation mixture by the
                     standard chemical method. The average insert size is about
                     3.2 kb. The library was constructed by Yulan Piao."
ORIGIN

Alignment Scores:
Pred. No.:                0.435          Length:          510
Score:                    79.00          Matches:         11
Percent Similarity:       100.00%        Conservative:    1
Best Local Similarity:    91.67%         Mismatches:      0
Query Match:              94.05%         Indels:          0
DB:                       6              Gaps:            0

US-09-858-332G-15 (1-12) x CA890073 (1-510)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||:::||||||||||||||||||||
Db        169 CACAATCACAATCACAGTCACAATCACAACCACAAT 134

RESULT 136
CK844809/c
LOCUS       CK844809                532 bp    mRNA    linear   EST 05-MAR-2004
DEFINITION  UI-R-AE0-xm-c-03-0-UI.s10 UI-R-AE0 Rattus norvegicus cDNA clone
            UI-R-AE0-xm-c-03-0-UI 3', mRNA sequence.
ACCESSION   CK844809
VERSION     CK844809.1  GI:45203716
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 532)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  PUBMED    8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road , 4156  MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Distribution information can be found at
            http://genome.uiowa.edu/distribution/rat.html
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .532
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-AE0-xm-c-03-0-UI"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="UI-R-AE0"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-AE0
                     library is a non-normalized library constructed from  15
                     dpc rat ventricle.  The tag is a string of 5 nucleotides
                     present  between the Not I site and the oligo-dT track.
                     The library was  constructed as described by Bonaldo,
                     Lennon and Soares,  Genome Research 6: 791-806, 1996.
                     Tissue provided by Jim Lin, Department of Biology,
                     University of Iowa.
                     TAG_TISSUE=ventricle at 15 dpc
                     TAG_LIB=UI-R-AE0
                     TAG_SEQ=GTGTC"
ORIGIN

Alignment Scores:
Pred. No.:                0.455          Length:          532
Score:                    79.00          Matches:         11
Percent Similarity:       100.00%        Conservative:    1
Best Local Similarity:    91.67%         Mismatches:      0
Query Match:              94.05%         Indels:          0
DB:                       7              Gaps:            0

US-09-858-332G-15 (1-12) x CK844809 (1-532)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||:::
Db        532 CACAACCACAACCACAACCACAACCACAATCATCAC 497

RESULT 137
BH142693/c
LOCUS       BH142693                562 bp    DNA     linear   GSS 16-AUG-2001
DEFINITION  TDGDR64TH cTOG Lycopersicon esculentum genomic clone cTOG24L8,
            genomic survey sequence.
ACCESSION   BH142693
VERSION     BH142693.1  GI:15195555
KEYWORDS    GSS.
SOURCE      Lycopersicon esculentum (Solanum lycopersicum)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 562)
```

```
  AUTHORS   van der Hoeven,R., Sun,H., Cho,J., Utterback,T., Ronning,C. and
            Tanksley,S.
  TITLE     Tomato Demethylated Genomic DNA Sequences
  JOURNAL   Unpublished (2001)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html
            tomato demethylated genomic DNA
            Insert Length: 1270   Std Error: 0.00
            Seq primer: M13F-R
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .562
                     /organism="Lycopersicon esculentum"
                     /mol_type="genomic DNA"
                     /cultivar="E6203"
                     /db_xref="taxon:4081"
                     /clone="cTOG24L8"
                     /tissue_type="young leaves"
                     /dev_stage="12-14 weeks post harvest"
                     /lab_host="E.coli JM109"
                     /clone_lib="cTOG"
                     /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                     XhoI; This library was made from short EcoRI digested
                     fragments of the genome of Lycopersicon esculentum ligated
                     into pBS (SK-). The fragments were  cloned into the
                     methylation restrictive E.coli strain JM109 with the
                     purpose of enriching the library for non-methylated DNA
                     fragments.  This procedure may enrich the pool of cloned
                     fragments in JM109 cells for sequences representing
                     expressed genes. Average insert size 1.27 kb."
ORIGIN

Alignment Scores:
Pred. No.:                0.482          Length:          562
Score:                    79.00          Matches:         11
Percent Similarity:       100.00%        Conservative:    1
Best Local Similarity:    91.67%         Mismatches:      0
Query Match:              94.05%         Indels:          0
DB:                       9              Gaps:            0

US-09-858-332G-15 (1-12) x BH142693 (1-562)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||:::|||||||||||||||||||
Db        232 CACAACCATAATCATAGTCATAATCATAATCATAAT 197

RESULT 138
CZ285358/c
LOCUS       CZ285358                635 bp    DNA     linear   GSS 01-JUL-2005
DEFINITION  cp40g07.f Candida parapsilosis Random Genomic Library Candida
            parapsilosis genomic clone cp40g07, genomic survey sequence.
ACCESSION   CZ285358
VERSION     CZ285358.1  GI:68456668
KEYWORDS    GSS.
SOURCE      Candida parapsilosis
  ORGANISM  Candida parapsilosis
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 635)
  AUTHORS   Logue,M.E., Wong,S., Wolfe,K.H. and Butler,G.
  TITLE     A genome sequence survey shows that the pathogenic yeast Candida
            parapsilosis has a defective MTLa1 allele at its mating type locus
  JOURNAL   Eukaryot. Cell 4 (6), 1009-1017 (2005)
   PUBMED   15947193
COMMENT     Contact: Logue M
            Department of Biochemistry, Conway Institute of Biomolecular and
            Biomedical Research
            University College Dublin
            Dublin 4, Ireland
            Tel: +353 1 7166885
            Fax: +353 1 2837211
            Email: mary.e.logue@ucd.ie
            Class: plasmid ends.
FEATURES             Location/Qualifiers
     source          1. .635
                     /organism="Candida parapsilosis"
                     /mol_type="genomic DNA"
                     /strain="CLIB214"
                     /db_xref="taxon:5480"
                     /clone="cp40g07"
                     /clone_lib="Candida parapsilosis Random Genomic Library"
ORIGIN

Alignment Scores:
Pred. No.:                0.547          Length:          635
Score:                    79.00          Matches:         11
Percent Similarity:       100.00%        Conservative:    1
Best Local Similarity:    91.67%         Mismatches:      0
Query Match:              94.05%         Indels:          0
DB:                       10             Gaps:            0

US-09-858-332G-15 (1-12) x CZ285358 (1-635)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||:::
Db        297 CACAACCACAACCACAACCACAACCACAACCACCAC 262

RESULT 139
AU268952
LOCUS       AU268952                645 bp    mRNA    linear   EST 26-APR-2004
DEFINITION  AU268952 VS Dictyostelium discoideum cDNA clone VSI605 3', mRNA
            sequence.
ACCESSION   AU268952
VERSION     AU268952.1  GI:20527750
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 645)
  AUTHORS   Urushihara,H., Morio,T., Saito,T., Kohara,Y., Koriki,E., Ochiai,H.,
            Maeda,M., Williams,J.G., Takeuchi,I. and Tanaka,Y.
  TITLE     Analyses of cDNAs from growth and slug stages of Dictyostelium
            discoideum
  JOURNAL   Nucleic Acids Res. 32 (5), 1647-1653 (2004)
   PUBMED   15010511
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .645
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="VSI605"
                     /sex="mat A"
                     /dev_stage="vegetative"
                     /clone_lib="VS"
ORIGIN

Alignment Scores:
Pred. No.:                0.556          Length:          645
Score:                    79.00          Matches:         11
Percent Similarity:       100.00%        Conservative:    1
Best Local Similarity:    91.67%         Mismatches:      0
Query Match:              94.05%         Indels:          0
DB:                       1              Gaps:            0

US-09-858-332G-15 (1-12) x AU268952 (1-645)
```

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||:::|||||||||||||||||||||||||
Db        572 CATAATCATGATCATAATCATAATCATAATCACAAT 607

RESULT 140
CX535222
LOCUS       CX535222                649 bp    mRNA    linear   EST 11-JAN-2005
DEFINITION  s13dNF95H04MJ044_388621 Methyl Jasmonate-Elicited Root Cell
            Suspension Culture Medicago truncatula cDNA, mRNA sequence.
ACCESSION   CX535222
VERSION     CX535222.1  GI:57503941
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 649)
  AUTHORS   Achnine,L., He,X.-Z., Scott,A.D., May,G.D. and Dixon,R.A.
  TITLE     Medicago truncatula Methyl Jasmonate-Elicited Root Cell Suspension
            Culture Expressed Sequence Tags from the Samuel Roberts Noble
            Foundation - Center for Medicago Genomics Research
  JOURNAL   Unpublished (2005)
COMMENT     Contact: May GD
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA
            Tel: 580 224 6650
            Fax: 580 224 6692
            Email: gdmay@noble.org.
FEATURES             Location/Qualifiers
     source          1. .649
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /db_xref="taxon:3880"
                     /tissue_type="Root cell suspension culture"
                     /dev_stage="Seven-day old culture"
                     /clone_lib="Methyl Jasmonate-Elicited Root Cell Suspension
                     Culture"
                     /note="Vector: Lambda Zap; Medicago truncatula root cell
                     suspension culture elicited with 500 um methyl jasmonate.
                     The cDNA was directionally ligated into the Uni-Zap XR
                     vector (Stratagene) and packaged using the Gigapack III
                     Gold packaging extracts. Phagemids containing cDNA inserts
                     were in vivo excised from the recombinant Uni-ZAP XR
                     vector using ExAssist helper phage and the E. coli strain
                     XL1-Blue MRF' (Stratagene). Excised plasmids were plated
                     using SOLR cells."
ORIGIN

Alignment Scores:
Pred. No.:                0.56          Length:        649
Score:                    79.00         Matches:       11
Percent Similarity:       100.00%       Conservative:  1
Best Local Similarity:    91.67%        Mismatches:    0
Query Match:              94.05%        Indels:        0
DB:                       8             Gaps:          0

US-09-858-332G-15 (1-12) x CX535222 (1-649)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||||||||||||||||||||||||||||||
Db        549 CATCATCATAATCATAATCATAATCATAATCATAAT 584

RESULT 141
CZ284612/c
LOCUS       CZ284612                674 bp    DNA     linear   GSS 01-JUL-2005
DEFINITION  cp36e12.r Candida parapsilosis Random Genomic Library Candida
            parapsilosis genomic clone cp36e12, genomic survey sequence.
ACCESSION   CZ284612
VERSION     CZ284612.1  GI:68455814
KEYWORDS    GSS.
SOURCE      Candida parapsilosis
  ORGANISM  Candida parapsilosis
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 674)
  AUTHORS   Logue,M.E., Wong,S., Wolfe,K.H. and Butler,G.
  TITLE     A genome sequence survey shows that the pathogenic yeast Candida
            parapsilosis has a defective MTLa1 allele at its mating type locus
  JOURNAL   Eukaryot. Cell 4 (6), 1009-1017 (2005)
  PUBMED    15947193
COMMENT     Contact: Logue M
            Department of Biochemistry, Conway Institute of Biomolecular and
            Biomedical Research
            University College Dublin
            Dublin 4, Ireland
            Tel: +353 1 7166885
            Fax: +353 1 2837211
            Email: mary.e.logue@ucd.ie
            Class: plasmid ends.
FEATURES             Location/Qualifiers
     source          1. .674
                     /organism="Candida parapsilosis"
                     /mol_type="genomic DNA"
                     /strain="CLIB214"
                     /db_xref="taxon:5480"
                     /clone="cp36e12"
                     /clone_lib="Candida parapsilosis Random Genomic Library"
ORIGIN

Alignment Scores:
Pred. No.:                0.582         Length:        674
Score:                    79.00         Matches:       11
Percent Similarity:       100.00%       Conservative:  1
Best Local Similarity:    91.67%        Mismatches:    0
Query Match:              94.05%        Indels:        0
DB:                       10            Gaps:          0

US-09-858-332G-15 (1-12) x CZ284612 (1-674)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||||||||||||||||||||||||||||||
Db        468 CACTCTCACAACCACAACCACAACCACAACCACAAC 433

RESULT 142
CZ288826
LOCUS       CZ288826                743 bp    DNA     linear   GSS 01-JUL-2005
DEFINITION  cp60d09.f Candida parapsilosis Random Genomic Library Candida
            parapsilosis genomic clone cp60d09, genomic survey sequence.
ACCESSION   CZ288826
VERSION     CZ288826.1  GI:68460136
KEYWORDS    GSS.
SOURCE      Candida parapsilosis
  ORGANISM  Candida parapsilosis
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 743)
  AUTHORS   Logue,M.E., Wong,S., Wolfe,K.H. and Butler,G.
  TITLE     A genome sequence survey shows that the pathogenic yeast Candida
            parapsilosis has a defective MTLa1 allele at its mating type locus
  JOURNAL   Eukaryot. Cell 4 (6), 1009-1017 (2005)
  PUBMED    15947193
COMMENT     Contact: Logue M
            Department of Biochemistry, Conway Institute of Biomolecular and
            Biomedical Research
            University College Dublin
            Dublin 4, Ireland
            Tel: +353 1 7166885
            Fax: +353 1 2837211
            Email: mary.e.logue@ucd.ie
            Class: plasmid ends.
FEATURES             Location/Qualifiers
     source          1. .743
```

```
                     /organism="Candida parapsilosis"
                     /mol_type="genomic DNA"
                     /strain="CLIB214"
                     /db_xref="taxon:5480"
                     /clone="cp60d09"
                     /clone_lib="Candida parapsilosis Random Genomic Library"
ORIGIN

Alignment Scores:
Pred. No.:                0.645          Length:         743
Score:                    79.00          Matches:        11
Percent Similarity:       100.00%        Conservative:   1
Best Local Similarity:    91.67%         Mismatches:     0
Query Match:              94.05%         Indels:         0
DB:                       10             Gaps:           0

US-09-858-332G-15 (1-12) x CZ288826 (1-743)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||||||||||||||||||||||||||||||
Db        441 CACTCTCACAACCACAACCACAACCACAACCACAAC 476

RESULT 143
BJ434857/c
LOCUS       BJ434857                 755 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ434857 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv25n03 3', mRNA sequence.
ACCESSION   BJ434857
VERSION     BJ434857.1  GI:19409579
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 755)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .755
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv25n03"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:                0.656          Length:         755
Score:                    79.00          Matches:        11
Percent Similarity:       100.00%        Conservative:   1
Best Local Similarity:    91.67%         Mismatches:     0
Query Match:              94.05%         Indels:         0
DB:                       3              Gaps:           0

US-09-858-332G-15 (1-12) x BJ434857 (1-755)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||:::||||||||||||
Db        201 CACAATCACAATCACAATCACGATCACAATCACAAT 166

RESULT 144
CZ288827/c
LOCUS       CZ288827                 756 bp    DNA     linear   GSS 01-JUL-2005
DEFINITION  cp60d09.r Candida parapsilosis Random Genomic Library Candida
            parapsilosis genomic clone cp60d09, genomic survey sequence.
ACCESSION   CZ288827
VERSION     CZ288827.1  GI:68460137
KEYWORDS    GSS.
SOURCE      Candida parapsilosis
  ORGANISM  Candida parapsilosis
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 756)
  AUTHORS   Logue,M.E., Wong,S., Wolfe,K.H. and Butler,G.
  TITLE     A genome sequence survey shows that the pathogenic yeast Candida
            parapsilosis has a defective MTLa1 allele at its mating type locus
  JOURNAL   Eukaryot. Cell 4 (6), 1009-1017 (2005)
  PUBMED    15947193
COMMENT     Contact: Logue M
            Department of Biochemistry, Conway Institute of Biomolecular and
            Biomedical Research
            University College Dublin
            Dublin 4, Ireland
            Tel: +353 1 7166885
            Fax: +353 1 2837211
            Email: mary.e.logue@ucd.ie
            Class: plasmid ends.
FEATURES             Location/Qualifiers
     source          1. .756
                     /organism="Candida parapsilosis"
                     /mol_type="genomic DNA"
                     /strain="CLIB214"
                     /db_xref="taxon:5480"
                     /clone="cp60d09"
                     /clone_lib="Candida parapsilosis Random Genomic Library"
ORIGIN

Alignment Scores:
Pred. No.:                0.656          Length:         756
Score:                    79.00          Matches:        11
Percent Similarity:       100.00%        Conservative:   1
Best Local Similarity:    91.67%         Mismatches:     0
Query Match:              94.05%         Indels:         0
DB:                       10             Gaps:           0

US-09-858-332G-15 (1-12) x CZ288827 (1-756)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||||||||||||||||||||||||||||||
Db        303 CACTCTCACAACCACAACCACAACCACAACCACAAC 268

RESULT 145
CZ288255/c
LOCUS       CZ288255                 783 bp    DNA     linear   GSS 01-JUL-2005
DEFINITION  cp57c02.r Candida parapsilosis Random Genomic Library Candida
            parapsilosis genomic clone cp57c02, genomic survey sequence.
ACCESSION   CZ288255
VERSION     CZ288255.1  GI:68459565
KEYWORDS    GSS.
SOURCE      Candida parapsilosis
  ORGANISM  Candida parapsilosis
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 783)
  AUTHORS   Logue,M.E., Wong,S., Wolfe,K.H. and Butler,G.
  TITLE     A genome sequence survey shows that the pathogenic yeast Candida
            parapsilosis has a defective MTLa1 allele at its mating type locus
  JOURNAL   Eukaryot. Cell 4 (6), 1009-1017 (2005)
  PUBMED    15947193
COMMENT     Contact: Logue M
            Department of Biochemistry, Conway Institute of Biomolecular and
            Biomedical Research
            University College Dublin
            Dublin 4, Ireland
            Tel: +353 1 7166885
```

```
           Fax: +353 1 2837211
           Email: mary.e.logue@ucd.ie
           Class: plasmid ends.
FEATURES             Location/Qualifiers
     source          1. .783
                     /organism="Candida parapsilosis"
                     /mol_type="genomic DNA"
                     /strain="CLIB214"
                     /db_xref="taxon:5480"
                     /clone="cp57c02"
                     /clone_lib="Candida parapsilosis Random Genomic Library"
ORIGIN

Alignment Scores:
Pred. No.:              0.681          Length:         783
Score:                  79.00          Matches:        11
Percent Similarity:     100.00%        Conservative:   1
Best Local Similarity:  91.67%         Mismatches:     0
Query Match:            94.05%         Indels:         0
DB:                     10             Gaps:           0

US-09-858-332G-15 (1-12) x CZ288255 (1-783)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||||||||||||||||||||||||||||||
Db        468 CACTCTCACAACCACAACCACAACCACAACCACAAC 433

RESULT 146
CC653644
LOCUS       CC653644                970 bp    DNA     linear   GSS 19-JUN-2003
DEFINITION  OGUIJ89TV ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMa0447O10,
            genomic survey sequence.
ACCESSION   CC653644
VERSION     CC653644.1  GI:32056611
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 970)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
            Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
  TITLE     Consortium for Maize Genomics
  JOURNAL   Unpublished (2002)
COMMENT     Other_GSSs: OGUIJ89TH
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: methylation filtered.
FEATURES             Location/Qualifiers
     source          1. .970
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBMa0447O10"
                     /clone_lib="ZM_0.7_1.5_KB"
                     /note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb
                     methylation filtered genomic DNA library"
ORIGIN

Alignment Scores:
Pred. No.:              0.852          Length:         970
Score:                  79.00          Matches:        11
Percent Similarity:     100.00%        Conservative:   1
Best Local Similarity:  91.67%         Mismatches:     0
Query Match:            94.05%         Indels:         0
DB:                     9              Gaps:           0

US-09-858-332G-15 (1-12) x CC653644 (1-970)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||||||||||||||||||||||||:::||||||
Db        583 CACAACCACAACCACAACCACAACCACAGCCACAAC 618

RESULT 147
CNS06LYT
LOCUS       CNS06LYT                1027 bp    DNA     linear   GSS 17-JUN-2001
DEFINITION  T7 end of clone XAT0AA002F04 of library XAT0AA from strain CBS 4311
            of Saccharomyces servazzii, genomic survey sequence.
ACCESSION   AL404811
VERSION     AL404811.1  GI:12166723
KEYWORDS    GSS.
SOURCE      Saccharomyces servazzii
  ORGANISM  Saccharomyces servazzii
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 1027)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  PUBMED    11152876
REFERENCE   2  (bases 1 to 1027)
  AUTHORS   Casaregola,S., Lepingle,A., Bon,E., Neuveglise,C., Nguyen,H.,
            Artiguenave,F., Wincker,P. and Gaillardin,C.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 7.
            Saccharomyces servazzii
  JOURNAL   FEBS Lett. 487 (1), 47-51 (2000)
  PUBMED    11152882
REFERENCE   3  (bases 1 to 1027)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .1027
                     /organism="Saccharomyces servazzii"
                     /mol_type="genomic DNA"
                     /strain="CBS 4311"
                     /db_xref="taxon:27293"
                     /clone="XAT0AA002F04"
                     /clone_lib="XAT0AA"
                     /note="end : T7"
     misc_feature    <349. .>1019
                     /note="similar to Saccharomyces cerevisiae ORF YBL097w [
                     BRN1 ; functional homologue of human BRRN1 ]
                     1 putative frameshift(s)"
                     /evidence=not_experimental
ORIGIN

Alignment Scores:
Pred. No.:              0.904          Length:         1027
Score:                  79.00          Matches:        11
```

```
Percent Similarity:     100.00%         Conservative:  1
Best Local Similarity:  91.67%          Mismatches:    0
Query Match:            94.05%          Indels:        0
DB:                     11              Gaps:          0

US-09-858-332G-15 (1-12) x CNS06LYT (1-1027)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||||||||||||||||||||||||||||||
Db        250 CATCATCACAACCACAACCACAACCACAACCATAAC 285

RESULT 148
CNS06LAO
LOCUS       CNS06LAO                1126 bp    DNA     linear   GSS 17-JUN-2001
DEFINITION  T7 end of clone AT0AA012A02 of library AT0AA from strain CBS 4311
            of Saccharomyces servazzii, genomic survey sequence.
ACCESSION   AL403942
VERSION     AL403942.1  GI:12164834
KEYWORDS    GSS.
SOURCE      Saccharomyces servazzii
  ORGANISM  Saccharomyces servazzii
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 1126)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  PUBMED    11152876
REFERENCE   2  (bases 1 to 1126)
  AUTHORS   Casaregola,S., Lepingle,A., Bon,E., Neuveglise,C., Nguyen,H.,
            Artiguenave,F., Wincker,P. and Gaillardin,C.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 7.
            Saccharomyces servazzii
  JOURNAL   FEBS Lett. 487 (1), 47-51 (2000)
  PUBMED    11152882
REFERENCE   3  (bases 1 to 1126)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .1126
                     /organism="Saccharomyces servazzii"
                     /mol_type="genomic DNA"
                     /strain="CBS 4311"
                     /db_xref="taxon:27293"
                     /clone="AT0AA012A02"
                     /clone_lib="AT0AA"
                     /note="end : T7"
     misc_feature    <445. .>1106
                     /note="similar to Saccharomyces cerevisiae ORF YBL097w [
                     BRN1 ; functional homologue of human BRRN1 ]
                     1 putative frameshift(s)"
                     /evidence=not_experimental
ORIGIN

Alignment Scores:
Pred. No.:              0.996           Length:        1126
Score:                  79.00           Matches:       11
Percent Similarity:     100.00%         Conservative:  1
Best Local Similarity:  91.67%          Mismatches:    0
Query Match:            94.05%          Indels:        0
DB:                     11              Gaps:          0

US-09-858-332G-15 (1-12) x CNS06LAO (1-1126)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||||||||||||||||||||||||||||||
Db        346 CATCATCACAACCACAACCACAACCACAACCATAAC 381

RESULT 149
CG895908/c
LOCUS       CG895908                 253 bp    DNA     linear   GSS 08-DEC-2003
DEFINITION  pastbac005xf03.b1.ab1 Res147 1 Pasteuria penetrans genomic 5',
            genomic survey sequence.
ACCESSION   CG895908
VERSION     CG895908.1  GI:39551403
KEYWORDS    GSS.
SOURCE      Pasteuria penetrans
  ORGANISM  Pasteuria penetrans
            Bacteria; Firmicutes; Bacillales; Alicyclobacillaceae; Pasteuria.
REFERENCE   1  (bases 1 to 253)
  AUTHORS   Opperman,C.H., Davies,K.G., Sosinski,B.R., Waterman,J. and Burke,M.
  TITLE     Unpublished Data
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Opperman CH
            Center for the Biology of Nematode Parasitism and Nematode
            Interactions Unit
            North Carolina State University and Rothamsted Research, Ltd.
            Box 7253, NCSU, Raleigh, NC 27606, USA
            Email: warthog@unity.ncsu.edu
            Homology: e-val = 1e-10. Description: None found.  row: f  column:
            3
            Class: shotgun
            High quality sequence stop: 253.
FEATURES             Location/Qualifiers
     source          1. .253
                     /organism="Pasteuria penetrans"
                     /mol_type="genomic DNA"
                     /strain="Res147"
                     /db_xref="taxon:86005"
                     /clone_lib="Res147 1"
ORIGIN

Alignment Scores:
Pred. No.:              0.281           Length:        253
Score:                  78.00           Matches:       11
Percent Similarity:     100.00%         Conservative:  0
Best Local Similarity:  100.00%         Mismatches:    0
Query Match:            92.86%          Indels:        0
DB:                     10              Gaps:          0

US-09-858-332G-15 (1-12) x CG895908 (1-253)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              ||||||||||||||||||||||||||||||||
Db        251 CACAACCACAACCACAACCACAACCACAACCAC 219

RESULT 150
CG898233/c
LOCUS       CG898233                 338 bp    DNA     linear   GSS 08-DEC-2003
DEFINITION  pastbac066xb17.b1.ab1 Res147 1 Pasteuria penetrans genomic 5',
            genomic survey sequence.
ACCESSION   CG898233
VERSION     CG898233.1  GI:39553728
KEYWORDS    GSS.
SOURCE      Pasteuria penetrans
  ORGANISM  Pasteuria penetrans
            Bacteria; Firmicutes; Bacillales; Alicyclobacillaceae; Pasteuria.
```

```
REFERENCE   1  (bases 1 to 338)
  AUTHORS   Opperman,C.H., Davies,K.G., Sosinski,B.R., Waterman,J. and Burke,M.
  TITLE     Unpublished Data
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Opperman CH
            Center for the Biology of Nematode Parasitism and Nematode
            Interactions Unit
            North Carolina State University and Rothamsted Research, Ltd.
            Box 7253, NCSU, Raleigh, NC 27606, USA
            Email: warthog@unity.ncsu.edu
            Homology: e-val = 1e-10. Description: None found.  row: b  column:
            17
            Class: shotgun
            High quality sequence stop: 338.
FEATURES             Location/Qualifiers
     source          1. .338
                     /organism="Pasteuria penetrans"
                     /mol_type="genomic DNA"
                     /strain="Res147"
                     /db_xref="taxon:86005"
                     /clone_lib="Res147 1"
ORIGIN

Alignment Scores:
Pred. No.:              0.381          Length:         338
Score:                  78.00          Matches:        11
Percent Similarity:     91.67%         Conservative:   0
Best Local Similarity:  91.67%         Mismatches:     1
Query Match:            92.86%         Indels:         0
DB:                     10             Gaps:           0

US-09-858-332G-15 (1-12) x CG898233 (1-338)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||||||||||||||||||||||||||   ||||||
Db        272 CACAACCACAACCACAACCACAACCACAAGCACAAC 237

RESULT 151
CV252723/c
LOCUS       CV252723                349 bp    mRNA    linear   EST 22-SEP-2004
DEFINITION  PX0015.B21_M04 PT-X-FL-A-1 Populus trichocarpa cDNA clone
            PX0015_M04 3', mRNA sequence.
ACCESSION   CV252723
VERSION     CV252723.1  GI:52505698
KEYWORDS    EST.
SOURCE      Populus trichocarpa (Populus balsamifera subsp. trichocarpa)
  ORGANISM  Populus trichocarpa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Malpighiales; Salicaceae; Saliceae; Populus.
REFERENCE   1  (bases 1 to 349)
  AUTHORS   Ralph,S., Cooper,D., Kolosova,N., Oddy,C., Butterfield,Y.,
            Kirkpatrick,R., Liu,J., Palmquist,D., Stott,J., Barber,S., Yang,G.,
            Babakaiff,R., Brown-John,M., Chand,S., Featherstone,R., Masson,A.,
            Mayo,M., Moran,J., Olson,T., Wong,D., Ritland,C.E., Siddiqui,A.,
            Holt,R., Jones,S., Marra,M., Ellis,B.E., Douglas,C., Ritland,K. and
            Bohlmann,J.
  TITLE     The poplar transcriptome: Analysis of expressed sequence tags from
            multiple cDNA libraries
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Joerg Bohlmann
            Genome BC forest genomics program
            University of British Columbia
            Michael Smith Laboratories, 6174 University Boulevard, Rm. 237,
            Vancouver, British Columbia, Canada, V6T 1Z3
            Tel: 1-604-822-0282
            Fax: 1-604-822-2114
            Email: bohlmann@msl.ubc.ca
            Plate: PX0015  row: M  column: 04
            High quality sequence stop: 349
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .349
                     /organism="Populus trichocarpa"
                     /mol_type="mRNA"
                     /cultivar="383-2499 (Nisqually-1)"
                     /db_xref="taxon:3694"
                     /clone="PX0015_M04"
                     /sex="Female"
                     /lab_host="E. coli DH10B T1 phage resistant cells"
                     /clone_lib="PT-X-FL-A-1"
                     /note="Vector: pBluescript II SK (+) XR; Site_1: SstI (5'
                     end of cDNA); Site_2: XhoI (3' end of cDNA); Outer xylem
                     from 8 year old trees harvested within the Boise Cascade
                     region of Washington state on May 15th, 2001. cDNA was
                     prepared from 20 micrograms of mRNA according to the
                     full-length cDNA library construction method described by
                     Carninci P. et al. (2000), Genome Research
                     10(10):1617-1630 and directionally ligated into the
                     pBluescript II SK (+) XR vector digested with SstI (5'
                     end) and XhoI (3'). Plasmid DNA was then transformed by
                     electroporation into DH10B cells (Invitrogen) for
                     propagation."
ORIGIN

Alignment Scores:
Pred. No.:              0.394          Length:         349
Score:                  78.00          Matches:        11
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            92.86%         Indels:         0
DB:                     7              Gaps:           0

US-09-858-332G-15 (1-12) x CV252723 (1-349)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              ||||||||||||||||||||||||||||||||||
Db        304 CACAACCACAACCACAACCACAACCACAACCAC 272

RESULT 152
BP866389/c
LOCUS       BP866389                376 bp    mRNA    linear   EST 17-FEB-2005
DEFINITION  BP866389 RAFL21 Arabidopsis thaliana cDNA clone RAFL21-77-I23 5',
            mRNA sequence.
ACCESSION   BP866389
VERSION     BP866389.1  GI:59927362
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 376)
  AUTHORS   Seki,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M., Akiyama,K.,
            Iida,K., Enju,A., Sakurai,T., Arakawa,T., Carninci,P., Fukuda,S.,
            Iida,J., Kawai,J., Sasaki,D., Shiraki,T., Hayashizaki,Y. and
            Shinozaki,K.
  TITLE     Large-Scale Analysis of RIKEN Arabidopsis Full-length cDNAs
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            An Arabidopsis full-length cDNA library was constructed essentially
            as reported previously (Seki et al., 1998, 2002).This clone is in a
            modified pBluescript vector.
            Please visit our web site (http://pfgweb.gsc.riken.jp and
            http://rarge.gsc.riken.jp) for further details.
FEATURES             Location/Qualifiers
     source          1. .376
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
```

```
                     /clone="RAFL21-77-I23"
                     /lab_host="DH10B"
                     /clone_lib="RAFL21"
                     /note="Site_1: BamHI; Site_2: SalI; Subtraction Library.
                     The sequence was obtained from samples subjected to
                     various stress and plant hormones-treated"
ORIGIN

Alignment Scores:
Pred. No.:                0.425          Length:         376
Score:                    78.00          Matches:        11
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              92.86%         Indels:         0
DB:                       3              Gaps:           0

US-09-858-332G-15 (1-12) x BP866389 (1-376)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              ||||||||||||||||||||||||||||||||||
Db        143 CACAATCACAATCACAATCACAATCACAATCAC 111

RESULT 153
BU017925/c
LOCUS       BU017925                 430 bp    mRNA    linear   EST 23-AUG-2002
DEFINITION  QHE17D20.yg.ab1 QH_EFGHJ sunflower RHA280 Helianthus annuus cDNA
            clone QHE17D20, mRNA sequence.
ACCESSION   BU017925
VERSION     BU017925.1  GI:22453445
KEYWORDS    EST.
SOURCE      Helianthus annuus (common sunflower)
  ORGANISM  Helianthus annuus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; campanulids; Asterales; Asteraceae; Asteroideae;
            Heliantheae; Helianthus.
REFERENCE   1  (bases 1 to 430)
  AUTHORS   Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J.,
            Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y.,
            Lai,Z., Church,S., Jackson,L. and Bradford,K.
  TITLE     Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
            singleton, see http://cgpdb.ucdavis.edu/ for details.
            Plate: QHE17  row: D  column: 20.
FEATURES             Location/Qualifiers
     source          1. .430
                     /organism="Helianthus annuus"
                     /mol_type="mRNA"
                     /cultivar="RHA280"
                     /db_xref="taxon:4232"
                     /clone="QHE17D20"
                     /lab_host="E.coli"
                     /clone_lib="QH_EFGHJ sunflower RHA280"
                     /note="Vector: pBRcDNASfiAB; The library was constructed
                     from 11 different sources of RNA from a single genotype.
                     Separate cDNAs were generated using primers that
                     incorporated unique 5' and 3' tags to distinguish each
                     source of RNA. cDNAs were then pooled, size-fractionated,
                     directionally cloned into a custom medium-copy vector and
                     transformations made with four size classes to minimize
                     size bias. Details of each source of RNA and library
                     construction can be obtained at http://cgpdb.ucdavis.edu/
                     TAG_TISSUE=chemical induction
                     TAG_LIB=QH_EFGHJ sunflower RHA280
                     TAG_SEQ=TGTAGCCGGG"
ORIGIN

Alignment Scores:
Pred. No.:                0.489          Length:         430
Score:                    78.00          Matches:        11
Percent Similarity:       100.00%        Conservative:   1
Best Local Similarity:    91.67%         Mismatches:     0
Query Match:              92.86%         Indels:         0
DB:                       5              Gaps:           0

US-09-858-332G-15 (1-12) x BU017925 (1-430)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||:::||||||||||||||||||||||
Db        117 CATAACCATAATTATAATCATAACCATAACCATAAC 82

RESULT 154
CR297245
LOCUS       CR297245                 511 bp    DNA     linear   GSS 19-NOV-2004
DEFINITION  mte1-14A4RM1 BAC end, cultivar Jemalong A17 of Medicago truncatula,
            genomic survey sequence.
ACCESSION   CR297245
VERSION     CR297245.1  GI:44692554
KEYWORDS    GSS.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 511)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
FEATURES             Location/Qualifiers
     source          1. .511
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /cultivar="Jemalong A17"
                     /db_xref="taxon:3880"
                     /clone_lib="MTE1"
                     /note="Vector: pIndigoBAC ; Site_1: EcoRI ; Site_2: EcoRI
                     ; Debelle F. and Chalhoub B.
                     Genoscope sequence ID : mte1-14A4RM1"
ORIGIN

Alignment Scores:
Pred. No.:                0.586          Length:         511
Score:                    78.00          Matches:        11
Percent Similarity:       91.67%         Conservative:   0
Best Local Similarity:    91.67%         Mismatches:     1
Query Match:              92.86%         Indels:         0
DB:                       11             Gaps:           0

US-09-858-332G-15 (1-12) x CR297245 (1-511)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||    ||||||||||||||||
Db        125 CATAACCATAACCATACTCATAATCATAATCATAAT 160

RESULT 155
AW736366
LOCUS       AW736366                 543 bp    mRNA    linear   EST 07-SEP-2000
DEFINITION  EST332285 KV3 Medicago truncatula cDNA clone pKV3-12G8, mRNA
            sequence.
ACCESSION   AW736366
VERSION     AW736366.1  GI:7643149
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
```

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.

REFERENCE 1 (bases 1 to 543)
AUTHORS VandenBosch,K., Hurt,J., Moore,J., Beremand,P., Peng,H., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.
TITLE ESTs from roots of Medicago truncatula after Rhizobium inoculation
JOURNAL Unpublished (1999)
COMMENT Contact: VandenBosch K
Department of Plant Biology
University of Minnesota
220 BioSci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA
Tel: 612 624 2755
Fax: 612 625 1738
Email: kvandenb@cbs.umn.edu
Texas A&M EST name:T255806e
TIGR sequence name:MTEAF40TK
More information is available at. .
'http://chrysie.tamu.edu/medicago'
Seq primer: SKmod (CTA gAA CTA gtg gAT CC).

FEATURES Location/Qualifiers
source 1. .543
/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="pKV3-12G8"
/tissue_type="Seedling roots"
/dev_stage="3 days post-inoculation with Sinorhizobium meliloti"
/lab_host="E. coli strain XLOLR"
/clone_lib="KV3"
/note="Vector: pBluescript SK -; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into the Unizap XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in XLOLR cells."

ORIGIN

Alignment Scores:

Pred. No.:	0.625	Length:	543
Score:	78.00	Matches:	11
Percent Similarity:	91.67%	Conservative:	0
Best Local Similarity:	91.67%	Mismatches:	1
Query Match:	92.86%	Indels:	0
DB:	1	Gaps:	0

US-09-858-332G-15 (1-12) x AW736366 (1-543)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||   ||||||||||||||||||
Db        176 CATAACCATAACCATACTCATAATCATAATCATAAT 211
```

RESULT 156
CN884225
LOCUS CN884225 573 bp mRNA linear EST 04-JUN-2004
DEFINITION 010815AASB006103HT (AASB) Royal Gala 10 DAFB fruit Malus x domestica cDNA clone AASB006103, mRNA sequence.
ACCESSION CN884225
VERSION CN884225.1 GI:48270467
KEYWORDS EST.
SOURCE Malus x domestica
ORGANISM Malus x domestica
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids I; Rosales; Rosaceae; Maloideae; Malus.
REFERENCE 1 (bases 1 to 573)
AUTHORS Beuning,L., Bowen,J., Crowhurst,R., Gleave,A., Janssen,B., McArtney,S., Newcomb,R., Ross,G., Snowden,K., Walton,E. and Yauk,Y.
TITLE HortResearch Apple EST Project
JOURNAL Unpublished (2004)
COMMENT Contact: Gleave,A.
Sequencing Facility
The Horticulture and Food Research Institute of New Zealand Ltd
120 Mt Albert Rd, Mt Albert, Auckland, New Zealand
Tel: 00 64 09 815 4200
Fax: 00 64 09 815 4201
Email: est@hortresearch.co.nz.

FEATURES Location/Qualifiers
source 1. .573
/organism="Malus x domestica"
/mol_type="mRNA"
/db_xref="taxon:3750"
/clone="AASB006103"
/tissue_type="Young fruit"
/dev_stage="10 days after full bloom"
/clone_lib="(AASB) Royal Gala 10 DAFB fruit"
/note="Vector: pBluescript SK(-); Library sequenced by Genesis Research & Development"

ORIGIN

Alignment Scores:

Pred. No.:	0.661	Length:	573
Score:	78.00	Matches:	11
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	92.86%	Indels:	0
DB:	7	Gaps:	0

US-09-858-332G-15 (1-12) x CN884225 (1-573)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              ||||||||||||||||||||||||||||||||||||
Db        485 CATAACCACAATCACAACCATAACCACAACCAC 517
```

RESULT 157
BJ325871
LOCUS BJ325871 663 bp mRNA linear EST 05-MAR-2002
DEFINITION BJ325871 Dictyostelium discoideum cDNA library, AF Dictyostelium discoideum cDNA clone dda2p18 5', mRNA sequence.
ACCESSION BJ325871
VERSION BJ325871.1 GI:19156001
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 663)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the aggregation stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES Location/Qualifiers
source 1. .663
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dda2p18"
/sex="mat A"
/dev_stage="Aggregation stage"
/clone_lib="Dictyostelium discoideum cDNA library, AF"

ORIGIN

Alignment Scores:

Pred. No.:	0.77	Length:	663

Score:	78.00	Matches:	11
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	92.86%	Indels:	0
DB:	3	Gaps:	0

US-09-858-332G-15 (1-12) x BJ325871 (1-663)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              ||||||||||||||||||||||||||||||||||
Db        551 CACAACCACAACCACAACCACAACCACAACCAC 583
```

RESULT 158
CB676683

```
LOCUS       CB676683                691 bp    mRNA    linear   EST 09-APR-2003
DEFINITION  OSJNEe13B01.f OSJNEe Oryza sativa (japonica cultivar-group) cDNA
            clone OSJNEe13B01 5', mRNA sequence.
ACCESSION   CB676683
VERSION     CB676683.1  GI:29680408
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 691)
  AUTHORS   Jantasuriyarat,C., Gowda,M., Haller,K., Hatfield,J., Lu,G.,
            Stahlberg,E., Zhou,B., Li,H., Kim,H., Yu,Y., Dean,R.A., Wing,R.A.,
            Soderlund,C. and Wang,G.L.
  TITLE     Large-scale identification of expressed sequence tags involved in
            rice and rice blast fungus interaction
  JOURNAL   Plant Physiol. 138 (1), 105-115 (2005)
   PUBMED   15888683
COMMENT     Contact: Rod Wing
            Arizona Genomics Institute
            University of Arizona
            Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
            85721-0088, USA
            Tel: 520 626 3967
            Fax: 520 621 9288
            Email: rwing@genome.arizona.edu
            PCR PRimers
            FORWARD: gta aaa cga cgg cca gtg
            BACKWARD: gga aac agc tat gac cat g
            Plate: 13  row: B  column: 01
            Seq primer: gta aaa cga cgg cca gtg.
FEATURES             Location/Qualifiers
     source          1. .691
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="OSJNEe13B01"
                     /tissue_type="Leaf"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
                     /clone_lib="OSJNEe"
                     /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
                     XhoI; 24 hrs after innoculation with Rice Blast (70-15)"
ORIGIN
```

Alignment Scores:

Pred. No.:	0.804	Length:	691
Score:	78.00	Matches:	11
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	92.86%	Indels:	0
DB:	6	Gaps:	0

US-09-858-332G-15 (1-12) x CB676683 (1-691)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              ||||||||||||||||||||||||||||||||||
```

```
Db        382 CACAATCACAACCACAACCACAATCACAACCAC 414
```

RESULT 159
CO363639

```
LOCUS       CO363639                694 bp    mRNA    linear   EST 29-JUN-2004
DEFINITION  RTK1_10_H12.g1_A029 Roots minus potassium Pinus taeda cDNA clone
            RTK1_10_H12_A029 5', mRNA sequence.
ACCESSION   CO363639
VERSION     CO363639.1  GI:49444956
KEYWORDS    EST.
SOURCE      Pinus taeda (loblolly pine)
  ORGANISM  Pinus taeda
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
REFERENCE   1  (bases 1 to 694)
  AUTHORS   Pratt,L., Cordonnier-Pratt,M.-M., Lorenz,W.W., Zimmermann,C. and
            Dean,J.F.D.
  TITLE     An EST database from potassium-deficient loblolly pine (Pinus
            taeda) roots
  JOURNAL   Unpublished (2004)
COMMENT     Other_ESTs: RTK1_10_H12.b1_A029
            Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            RNA prepared and library constructed by W. Walter Lorenz (School of
            Forest Resources, University of Georgia); plant material prepared
            by Craig Zimmermann (School of Forest Resources, University of
            Georgia) using rooted cuttings provided by the Forest Biology
            Research Cooperative (FBRC) and the CCLONES project a the
            University of Florida; sequencing done in the Laboratory for
            Genomics and Bioinformatics, University of Georgia. Sequence ends
            have been trimmed to exclude vector and regions below Phred quality
            16. Three-prime sequences are presented as their reverse complement
            and have been trimmed to exclude polyA.
            Seq primer: JENREV (CAGGAAACAGCTATGACC).
FEATURES             Location/Qualifiers
     source          1. .694
                     /organism="Pinus taeda"
                     /mol_type="mRNA"
                     /strain="3 CCLONES"
                     /db_xref="taxon:3352"
                     /clone="RTK1_10_H12_A029"
                     /lab_host="DH10B-T1 phage-resistant E. coli"
                     /clone_lib="Roots minus potassium"
                     /note="Organ: Root; Vector: pSL1180; Site_1: EcoRI;
                     Site_2: XhoI; The library was prepared from polyA+ RNA
                     from the roots of 1-year-old loblolly pine (Pinus taeda)
                     cuttings that were rooted and then planted in washed sand.
                     The rooted cuttings were maintained for 117 days (July
                     2003 harvest) under ambient conditions in a local
                     greenhouse. They were kept on a weekly regimen of 0.5x
                     nutrient-complete Hoagland's solution and supplemented
                     with additional water sufficient to maintain a 15% soil
                     moisture content. For twenty-eight days (28 d) prior to
                     harvesting roots for mRNA preparation, the trees received
                     Hoagland's solution lacking potassium (K) to induce a
                     potassium-deficiency. Double-stranded cDNA was cloned
                     unidirectionally into pSL1180. Inserts can be excised with
                     EcoRI (5' end) and XhoI (3' end)."
ORIGIN
```

Alignment Scores:

Pred. No.:	0.807	Length:	694
Score:	78.00	Matches:	11
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	92.86%	Indels:	0
DB:	7	Gaps:	0

```
US-09-858-332G-15 (1-12) x CO363639 (1-694)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              ||||||||||||||||||||||||||||||||||
Db         58 CATAATCACAATCACAATCACAACCACAACCAT 90

RESULT 160
DR693890
LOCUS       DR693890                795 bp    mRNA    linear   EST 13-JUL-2005
DEFINITION  EST1083980 Normalized pine embryo library, Lib_D Pinus taeda cDNA
            clone PWAE553 3' end, mRNA sequence.
ACCESSION   DR693890
VERSION     DR693890.1  GI:70784242
KEYWORDS    EST.
SOURCE      Pinus taeda (loblolly pine)
  ORGANISM  Pinus taeda
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
REFERENCE   1  (bases 1 to 795)
  AUTHORS   Buell,C.R., Zheng,L., Cowles,A. and Cairney,J.
  TITLE     Sequencing of ESTs from loblolly pine embryonic libraries
  JOURNAL   Unpublished (2004)
COMMENT     Contact: C. Robin Buell
            Plant Genomics Group
            The Institute for Genomic Research
            9712 Medical Center Dr, Rockville, MD 20850, USA
            Email: rbuell@tigr.org
            This clone is available through TIGR. Please contact pine@tigr.org
            for further information
            Seq primer: TAA TAC GAC TCA CTA TAG GG.
FEATURES             Location/Qualifiers
     source          1. .795
                     /organism="Pinus taeda"
                     /mol_type="mRNA"
                     /cultivar="7-56 mother tree, open-pollinated tree from,
                     Lyons, Georgia, USA"
                     /db_xref="taxon:3352"
                     /clone="PWAE553"
                     /tissue_type="Whole embryos excised from megagametophytes"
                     /lab_host="DH10B"
                     /clone_lib="Normalized pine embryo library, Lib_D"
                     /note="Organ: Zygotic Embryos; Vector: pGADT7-RecAB;
                     Site_1: EcoRI; Site_2: ClaI; Library enriched in
                     transcripts present in suspension culture of Loblolly Pine
                     (Pinus taeda L.), 7-56. Suspension culture was established
                     from fertilized megagametophytes, according to method of
                     Pullman et al 2003 (Pullman, G., S. Johnson, G. Peter, J.
                     Cairney, and N. Xu. 2003. Improving loblolly pine somatic
                     embryo maturation: Comparison of somatic and zygotic
                     embryo morphology, germination, and gene expression. Plant
                     Cell Reports 21:747-758. ) This suspension cell culture,
                     which contains early developing embryos and
                     undifferentiated embryogeneic tissue, was the source of
                     the RNA used for the suspension culture library.
                     Subtracted from this was cDNA made from a pool of RNA from
                     zygotic embryos and female gametophytes from all stages of
                     development plus somatic embryos from developmental stages
                     beyond susepension culture (see Ciavatta VT, Morillon R,
                     Pullman GS, Chrispeels M, Cairney J. 2001. An
                     aquaglyceroporin is abundantly expressed early in the
                     development of the suspensor and the embryo proper of
                     loblolly pine (Pinus taeda L.). Plant Physiol. 127:
                     1556-1567
                     [http://www.plantphysiol.org/cgi/content/full/127/4/1556])
                     ."
ORIGIN

Alignment Scores:
Pred. No.:              0.931          Length:         795
Score:                  78.00          Matches:        11
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            92.86%         Indels:         0
DB:                     8              Gaps:           0

US-09-858-332G-15 (1-12) x DR693890 (1-795)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              ||||||||||||||||||||||||||||||||||
Db        630 CATAATCACAATCACAATCACAACCACAACCAT 662

RESULT 161
DR631262
LOCUS       DR631262                851 bp    mRNA    linear   EST 11-JUL-2005
DEFINITION  EST1021390 FvI Gibberella moniliformis cDNA clone FVIE256, mRNA
            sequence.
ACCESSION   DR631262
VERSION     DR631262.1  GI:70706102
KEYWORDS    EST.
SOURCE      Gibberella moniliformis
  ORGANISM  Gibberella moniliformis
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
REFERENCE   1  (bases 1 to 851)
  AUTHORS   Brown,D.W., Cheung,F., Proctor,R.H., Butchko,A.E., Zheng,L, Lee,Y.,
            Utterback,T., Smith,S., Feldblyum,T., Glenn,A.E., Plattner,R.D.,
            Kendra,D.F., Town,C.D. and Whitelaw,C.A.
  TITLE     Analysis of 87,000 expressed sequence tags reveals alternatively
            spliced introns in multiple genes of the fumonisin gene cluster
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Brown, D.W.
            USDA/ARS/NCAUR
            USDA
            1815 N. University St, Peoria, IL 61604, USA
            Tel: 309 681 6230
            Fax: 309 681 6689
            Email: browndw@ncaur.usda.gov
            TIGR sequence name: FVIE256TH
            Seq primer: AAT TAA CCC TCA CTA AAG GG.
FEATURES             Location/Qualifiers
     source          1. .851
                     /organism="Gibberella moniliformis"
                     /mol_type="mRNA"
                     /strain="m3125"
                     /db_xref="taxon:117187"
                     /clone="FVIE256"
                     /tissue_type="mycelia"
                     /clone_lib="FvI"
                     /note="Vector: pBlueScript II SK(+) XR; Site_1: EcoRI;
                     Site_2: XhoI; anamorph: Fusarium verticillioides. Library
                     FvI was prepared from growth on excised maize seedling
                     roots and shoots. The roots and shoots were prepared as
                     described above and then inoculated by dipping briefly in
                     a suspension of 5 x 10e6 conidia per ml. The inoculated
                     roots and shoots were then incubated on moistened Whatman
                     #1 filter paper at room temperature under sterile
                     conditions for 4 days. Fungal mycelial was collected,
                     frozen in liquid nitrogen, ground to a powder, and then
                     added to TRIzol Reagent (Invitrogen, Carlsbad CA) at
                     approximately 1 g mycelia per 10 ml TRIzol. The cDNA was
                     directionally ligated into the pBlueScript II SK(+) XR
                     vector (cDNA Synthesis Kit; Stratagene)."
ORIGIN

Alignment Scores:
Pred. No.:              0.999          Length:         851
Score:                  78.00          Matches:        11
Percent Similarity:     91.67%         Conservative:   0
Best Local Similarity:  91.67%         Mismatches:     1
Query Match:            92.86%         Indels:         0
DB:                     8              Gaps:           0

US-09-858-332G-15 (1-12) x DR631262 (1-851)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||   ||||||||||||||||||||||||||||||
```

```
Db         95 CACACCCACAACCACAACCACAACCACAACCACAAC 130

RESULT 162
BG298931
LOCUS       BG298931                1125 bp    mRNA    linear   EST 21-FEB-2001
DEFINITION  602397468F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4512065 5',
            mRNA sequence.
ACCESSION   BG298931
VERSION     BG298931.1  GI:13064079
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1125)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10396  row: h  column: 18
            High quality sequence stop: 552.
FEATURES             Location/Qualifiers
     source          1. .1125
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4512065"
                     /tissue_type="retina"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_94"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 3.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
ORIGIN

Alignment Scores:
Pred. No.:              1.34          Length:        1125
Score:                  78.00         Matches:       11
Percent Similarity:     91.67%        Conservative:  0
Best Local Similarity:  91.67%        Mismatches:    1
Query Match:            92.86%        Indels:        0
DB:                     2             Gaps:          0

US-09-858-332G-15 (1-12) x BG298931 (1-1125)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||   ||||||||||||||||||
Db        350 CACAACCACAACCACAAGCACAACCACAACCACAAC 385

RESULT 163
CB606620
LOCUS       CB606620                250 bp    mRNA    linear   EST 16-MAY-2003
DEFINITION  AMGNNUC:CR1-00147-A9-B Colon Rat 1 (10390) Rattus norvegicus cDNA
            clone cr1-00147-a9 3', mRNA sequence.
ACCESSION   CB606620
VERSION     CB606620.1  GI:29546233
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 250)
  AUTHORS   Amgen EST Program.
  TITLE     Amgen Rat EST Program
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Dan Fitzpatrick
            Amgen, Inc
            One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA
            Tel: 805 447-4881
            Plate: 00147  row: a  column: 9.
FEATURES             Location/Qualifiers
     source          1. .250
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
                     /clone="cr1-00147-a9"
                     /tissue_type="intestine, fetal"
                     /clone_lib="Colon Rat 1 (10390)"
                     /note="Vector: pMOB; Site_1: SalI; Site_2: NotI; Colon Rat
                     1 Fetal rat intestinal cDNA library."
ORIGIN

Alignment Scores:
Pred. No.:              0.373         Length:        250
Score:                  77.00         Matches:       11
Percent Similarity:     91.67%        Conservative:  0
Best Local Similarity:  91.67%        Mismatches:    1
Query Match:            91.67%        Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-15 (1-12) x CB606620 (1-250)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||   ||||||||||||
Db         84 CACAACCACAACCACAACCACANCCACAACCACAAC 119

RESULT 164
BJ351548/c
LOCUS       BJ351548                314 bp    mRNA    linear   EST 06-MAR-2002
DEFINITION  BJ351548 Dictyostelium discoideum cDNA library, AF Dictyostelium
            discoideum cDNA clone dda43h24 3', mRNA sequence.
ACCESSION   BJ351548
VERSION     BJ351548.1  GI:19222094
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 314)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the aggregation
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .314
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dda43h24"
                     /sex="mat A"
                     /dev_stage="Aggregation stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN

Alignment Scores:
Pred. No.:              0.474         Length:        314
Score:                  77.00         Matches:       11
```

```
Percent Similarity:     100.00%        Conservative:  1
Best Local Similarity:  91.67%         Mismatches:    0
Query Match:            91.67%         Indels:        0
DB:                     3              Gaps:          0

US-09-858-332G-15 (1-12) x BJ351548 (1-314)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              :::||||||||||||||||||||||||||||||||
Db        275 AACAACCACAACCACAACCACAACCACAACCACAAC 240

RESULT 165
BJ437953/c
LOCUS       BJ437953                327 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ437953 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv35k23 3', mRNA sequence.
ACCESSION   BJ437953
VERSION     BJ437953.1  GI:19412675
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 327)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .327
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv35k23"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:              0.494          Length:        327
Score:                  77.00          Matches:       11
Percent Similarity:     100.00%        Conservative:  1
Best Local Similarity:  91.67%         Mismatches:    0
Query Match:            91.67%         Indels:        0
DB:                     3              Gaps:          0

US-09-858-332G-15 (1-12) x BJ437953 (1-327)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              :::||||||||||||||||||||||||||||||||
Db        262 AACAACCACAACCACAACCACAACCACAACCACAAC 227

RESULT 166
BG362662
LOCUS       BG362662                519 bp    mRNA    linear   EST 22-JUL-2004
DEFINITION  sac07a11.y1 Gm-c1040 Glycine max cDNA clone GENOME SYSTEMS CLONE
            ID: Gm-c1040-3861 5' similar to TR:O23485 O23485 DNA CHROMOSOME 4,
            ESSA I CONTIG FRAGMENT NO. 6. ;, mRNA sequence.
ACCESSION   BG362662
VERSION     BG362662.1  GI:13251759
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 519)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,
            Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
            Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
            Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
            Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            When it has been determined, an EST from the other end of this
            clone is listed in the 'Other ESTs on clone' field. This clone is
            available through: Biogenetic Services, 801 32nd Ave. Brookings, SD
            57006 USA (phone: 800 423 4163; email: info@biogeneticservices.com)
            High quality sequence stop: 396.
FEATURES             Location/Qualifiers
     source          1. .519
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /cultivar="Williams 82"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1040-3861"
                     /tissue_type="Hypocotyl and Plumule, germinating seeds"
                     /lab_host="DH10B"
                     /clone_lib="Gm-c1040"
                     /note="Vector: pT7T3Pac (Pharmacia); Site_1: EcoRI;
                     Site_2: NotI; This cDNA library was constructed from mRNA
                     isolated from hypocotyl and plumule tissues of seeds
                     germinated for three days of the cultivar Williams 82.
                     Complementary DNA was synthesized from mRNA using a primer
                     consisting  of a poly(dT) sequence with a NotI restriction
                     site. EcoRI adapters were  ligated to the blunt-ended cDNA
                     fragments followed by digestions with EcoRI  and NotI. The
                     cDNA fragments were directionally cloned into the
                     EcoRI-NotI restriction site of the pT7T3-Pac vector. The
                     ligated cDNA fragments were transformed into DH10B host
                     cells (Gibco BRL). This library was constructed by Dr.
                     Randy Shoemaker."
ORIGIN

Alignment Scores:
Pred. No.:              0.801          Length:        519
Score:                  77.00          Matches:       11
Percent Similarity:     100.00%        Conservative:  1
Best Local Similarity:  91.67%         Mismatches:    0
Query Match:            91.67%         Indels:        0
DB:                     2              Gaps:          0

US-09-858-332G-15 (1-12) x BG362662 (1-519)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||:::|||
Db         11 CACAATCACAATCACAATCACAATCACAATAACAAT 46

RESULT 167
CA899924/c
LOCUS       CA899924                520 bp    mRNA    linear   EST 27-DEC-2002
DEFINITION  PCS00167M13RF Scarlet Runner Bean Suspensor Region pCR2.1 Phaseolus
            coccineus cDNA 5' similar to hypothetical protein, mRNA sequence.
ACCESSION   CA899924
VERSION     CA899924.1  GI:27386916
KEYWORDS    EST.
SOURCE      Phaseolus coccineus
  ORGANISM  Phaseolus coccineus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
```

```
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Phaseolus.
REFERENCE   1  (bases 1 to 520)
  AUTHORS   Bui,A.Q., Le,B.H., Weterings,K., Bi,Y.-P., Choi,J.-S.,
            McElroy,K.E., Choi,P.S., Harada,J.J., Fischer,R.L. and
            Goldberg,R.B.
  TITLE     Gene Activity in Different Regions of a Post-Fertilization Plant
            Embryo by EST Analysis
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Goldberg, R.B.
            Department of Molecular, Cell, & Developmental Biology
            University of California, Los Angeles
            621 Charles E. Young Drive South, Los Angeles, CA 90095-1606, USA
            Tel: 310 825 3270
            Fax: 310 825 8201
            Email: bobg@ucla.edu
            Seq primer: M13R
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .520
                     /organism="Phaseolus coccineus"
                     /mol_type="mRNA"
                     /cultivar="Hammond's Dwarf Scarlet"
                     /db_xref="taxon:3886"
                     /dev_stage="6-days post-pollination"
                     /clone_lib="Scarlet Runner Bean Suspensor Region pCR2.1"
                     /note="Organ: Suspensor Region of Globular-Stage Embryos;
                     Vector: pCR2.1; Site_1: EcoRI; Site_2: EcoRI; Suspensor
                     regions were micro-dissected from globular-stage embryos
                     six days after pollination from greenhouse-grown plants
                     [Weterings et al., Plant Cell 13, 2409-2425 (2001)].
                     Double-stranded cDNA was synthesized from suspensor mRNA
                     using the SMART cDNA Library Construction Kit according to
                     the manufacturer (Clontech). The suspensor cDNA fragments
                     were ligated non-directionally into the pCR2.1 vector
                     (Invitrogen), and the recombinant cDNAs were transformed
                     into E. coli INV alpha F' cells (Invitrogen)."
ORIGIN

Alignment Scores:
Pred. No.:                0.803          Length:          520
Score:                    77.00          Matches:         11
Percent Similarity:       100.00%        Conservative:    1
Best Local Similarity:    91.67%         Mismatches:      0
Query Match:              91.67%         Indels:          0
DB:                       6              Gaps:            0

US-09-858-332G-15 (1-12) x CA899924 (1-520)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||:::|||||||||||||||
Db        101 CACAATCACAACCACAACAACAACCACAACCACAAC 66

RESULT 168
AU053519
LOCUS       AU053519                 531 bp    mRNA    linear   EST 28-APR-1999
DEFINITION  AU053519 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
            discoideum cDNA clone SLI829, mRNA sequence.
ACCESSION   AU053519
VERSION     AU053519.1  GI:4702001
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 531)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp
            PROJECT = Dictyostelium discoideum cDNA project in Japan.
FEATURES             Location/Qualifiers
     source          1. .531
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SLI829"
                     /dev_stage="slug"
                     /clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
ORIGIN

Alignment Scores:
Pred. No.:                0.821          Length:          531
Score:                    77.00          Matches:         11
Percent Similarity:       100.00%        Conservative:    1
Best Local Similarity:    91.67%         Mismatches:      0
Query Match:              91.67%         Indels:          0
DB:                       1              Gaps:            0

US-09-858-332G-15 (1-12) x AU053519 (1-531)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              :::||||||||||||||||||||||||||||||||
Db        187 AACAACCACAACCACAACCACAACCACAATCACAAC 222

RESULT 169
BJ384768/c
LOCUS       BJ384768                 717 bp    mRNA    linear   EST 08-MAR-2002
DEFINITION  BJ384768 Dictyostelium discoideum cDNA library, CF Dictyostelium
            discoideum cDNA clone ddc50113 3', mRNA sequence.
ACCESSION   BJ384768
VERSION     BJ384768.1  GI:19294152
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 717)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the culmination
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .717
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddc50113"
                     /sex="mat A"
                     /dev_stage="Culmination stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, CF"
ORIGIN

Alignment Scores:
Pred. No.:                1.12           Length:          717
Score:                    77.00          Matches:         11
Percent Similarity:       100.00%        Conservative:    1
Best Local Similarity:    91.67%         Mismatches:      0
Query Match:              91.67%         Indels:          0
DB:                       3              Gaps:            0
```

```
US-09-858-332G-15 (1-12) x BJ384768 (1-717)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||:::|||||||||
Db        266 CACAATCACAACCACAACCACAACAACAACCACAAC 231

RESULT 170
BU994168
LOCUS       BU994168                247 bp    mRNA    linear   EST 23-OCT-2002
DEFINITION  HM05P14r HM Hordeum vulgare subsp. vulgare cDNA clone HM05P14
            5-PRIME, mRNA sequence.
ACCESSION   BU994168
VERSION     BU994168.1  GI:24271151
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 247)
  AUTHORS   Zhang,H., Weschke,W., Michalek,W., Stein,N. and Graner,A.
  TITLE     EST sequencing and analysis in barley (2002)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Stein Nils
            Molecular Markers Group, Department Genbank
            Institute of Plant Genetics and Crop Plant Research (IPK)
            Correnssstr. 3, 06466, Gatersleben, Germany
            Tel: 039482-5522
            Fax: 039482-5595
            Email: stein@ipk-gatersleben.de
            Insert Length: 247   Std Error: 0.00
            Plate: 5  row: P  column: 14
            Seq primer: M13rev.
FEATURES             Location/Qualifiers
     source          1. .247
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Barke"
                     /sub_species="vulgare"
                     /db_xref="GABI:253954"
                     /db_xref="taxon:112509"
                     /clone="HM05P14"
                     /tissue_type="male inflorescences"
                     /dev_stage="male inflorescences (approx. 2 mm in size),
                     green anther stage"
                     /lab_host="XL10-Gold"
                     /clone_lib="HM"
                     /note="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of
                     cDNA); Site_2: XhoI (3'-end of cDNA); Due to a cloning
                     artefact caused by the kit, in most cases the EcoRI site
                     is NOT present, as well as the EcoRIadapter used for
                     cloning. To excise the insert, restriction sites upstream
                     EcoRI should be used (e.g. BamHI, SalI,PstI). NOTE: Also
                     due to the cloning system used Blue/white selection for
                     recombinats is not 100% reliable."
ORIGIN

Alignment Scores:
Pred. No.:              0.496          Length:         247
Score:                  76.00          Matches:        11
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            90.48%         Indels:         0
DB:                     5              Gaps:           0

US-09-858-332G-15 (1-12) x BU994168 (1-247)

Qy          2 AsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||
Db        117 AACCACAACCACAACCACAACCACAACCACAAC 149

RESULT 171
BQ470935
LOCUS       BQ470935                334 bp    mRNA    linear   EST 30-MAY-2002
DEFINITION  HX04H16r HX Hordeum vulgare subsp. vulgare cDNA clone HX04H16
            5-PRIME, mRNA sequence.
ACCESSION   BQ470935
VERSION     BQ470935.1  GI:21278717
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 334)
  AUTHORS   Zhang,H., Weschke,W., Michalek,W., Stein,N. and Graner,A.
  TITLE     EST sequencing and analysis in barley (2002)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Stein Nils
            Molecular Markers Group, Department Genbank
            Institute of Plant Genetics and Crop Plant Research (IPK)
            Correnssstr. 3, 06466, Gatersleben, Germany
            Tel: 039482-5522
            Fax: 039482-5595
            Email: stein@ipk-gatersleben.de
            Insert Length: 334   Std Error: 0.00
            Plate: 4  row: H  column: 16
            Seq primer: M13rev.
FEATURES             Location/Qualifiers
     source          1. .334
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Barke"
                     /sub_species="vulgare"
                     /db_xref="taxon:112509"
                     /clone="HX04H16"
                     /tissue_type="apex"
                     /dev_stage="apex (3-5 mm in size)"
                     /lab_host="XL10-Gold"
                     /clone_lib="HX"
                     /note="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of
                     cDNA); Site_2: XhoI (3'-end of cDNA); Due to a cloning
                     artefact caused by the kit, in most cases the EcoRI site
                     is NOT present, as well as the EcoRIadapter used for
                     cloning. To excise the insert, restriction sites upstream
                     EcoRI should be used (e.g. BamHI, SalI,PstI). NOTE: Also
                     due to the cloning system used Blue/white selection for
                     recombinats is not 100% reliable."
ORIGIN

Alignment Scores:
Pred. No.:              0.68           Length:         334
Score:                  76.00          Matches:        11
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            90.48%         Indels:         0
DB:                     5              Gaps:           0

US-09-858-332G-15 (1-12) x BQ470935 (1-334)

Qy          2 AsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||
Db        245 AACCACAACCACAACCACAACCACAACCACAAC 277

RESULT 172
AW332717/c
LOCUS       AW332717                390 bp    mRNA    linear   EST 31-JAN-2000
DEFINITION  S12A5 AGS-1 Pneumocystis carinii cDNA 3', mRNA sequence.
ACCESSION   AW332717
VERSION     AW332717.1  GI:6829074
KEYWORDS    EST.
SOURCE      Pneumocystis carinii
  ORGANISM  Pneumocystis carinii
            Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;
            Pneumocystidaceae; Pneumocystis.
```

```
REFERENCE   1  (bases 1 to 390)
  AUTHORS   Smulian,A.G., Arnold,J., Weise,M., Wunderlich,J., Staben,C.,
            Edman,J.C., Kovacs,J. and Cushion,M.
  TITLE     Expressed sequence tags from Pneumocystis carinii
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Staben C
            School of Biological Sciences
            University of Kentucky
            101 Morgan Building, University of Kentucky, Lexington, KY
            40506-0225, USA
            Tel: 606 257 2161
            Fax: 606 257 1717
            Email: staben@pop.uky.edu.
FEATURES             Location/Qualifiers
     source          1. .390
                     /organism="Pneumocystis carinii"
                     /mol_type="mRNA"
                     /db_xref="taxon:4754"
                     /lab_host="E. coli"
                     /clone_lib="AGS-1"
                     /note="Vector: Lambda ZAP II; Site_1: EcoRI; Site_2: XhoI;
                     P. carinii organisms (3x10e9) from a single rat (99-1-6,
                     sacrificed on 3/17/99)  at Cincinnati VA facilities.
                     Trizol extracted RNA.  Oligo dT priming, standard
                     conditions described by  vendor, Stratagene. Further
                     details see www.uky.edu/Project/Pneumocystis/"
ORIGIN

Alignment Scores:
Pred. No.:                0.799          Length:         390
Score:                    76.00          Matches:        11
Percent Similarity:       91.67%         Conservative:   0
Best Local Similarity:    91.67%         Mismatches:     1
Query Match:              90.48%         Indels:         0
DB:                       1              Gaps:           0

US-09-858-332G-15 (1-12) x AW332717 (1-390)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||   |||
Db        357 CACAATCACAATCACAATCACAATCACAATCGCAAC 322

RESULT 173
CA023404
LOCUS       CA023404                516 bp    mRNA    linear   EST 23-OCT-2002
DEFINITION  HZ46C09r HZ Hordeum vulgare subsp. vulgare cDNA clone HZ46C09
            5-PRIME, mRNA sequence.
ACCESSION   CA023404
VERSION     CA023404.1  GI:24300778
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 516)
  AUTHORS   Radchuk,V., Zhang,H., Weschke,W., Potokina,E. and Wobus,U.
  TITLE     Barley ESTs from developing seeds
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Stein Nils
            Molecular Markers Group, Department Genbank
            Institute of Plant Genetics and Crop Plant Research (IPK)
            Correnssstr. 3, 06466, Gatersleben, Germany
            Tel: 039482-5522
            Fax: 039482-5595
            Email: stein@ipk-gatersleben.de
            Insert Length: 516   Std Error: 0.00
            Plate: 46  row: C  column: 9
            Seq primer: M13rev.
FEATURES             Location/Qualifiers
     source          1. .516
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Barke"
                     /sub_species="vulgare"
                     /db_xref="GABI:276813"
                     /db_xref="taxon:112509"
                     /clone="HZ46C09"
                     /tissue_type="pericarp"
                     /dev_stage="0-7 DAP (days after pollination)"
                     /lab_host="XL10-Gold"
                     /clone_lib="HZ"
                     /note="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of
                     cDNA); Site_2: XhoI (3'-end of cDNA); pericarp 0-7
                     DAP(days after pollination). Due to a cloning artefact
                     caused by the kit, in most cases the EcoRI site is NOT
                     present, as well as the EcoRIadapter used for cloning. To
                     excise the insert, restriction sites upstream EcoRI should
                     be used (e.g. BamHI, SalI,PstI). NOTE: Also due to the
                     cloning system used Blue/white selection for recombinats
                     is not 100% reliable.Average insert size is 900 bp"
ORIGIN

Alignment Scores:
Pred. No.:                1.07           Length:         516
Score:                    76.00          Matches:        11
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              90.48%         Indels:         0
DB:                       5              Gaps:           0

US-09-858-332G-15 (1-12) x CA023404 (1-516)

Qy          2 AsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||
Db        176 AACCACAACCACAACCACAACCACAACCACAAC 208

RESULT 174
BF254321
LOCUS       BF254321                527 bp    mRNA    linear   EST 22-OCT-2001
DEFINITION  HVSMEf0003K07f Hordeum vulgare seedling root EST library HVcDNA0007
            (Etiolated and unstressed) Hordeum vulgare subsp. vulgare cDNA
            clone HVSMEf0003K07f, mRNA sequence.
ACCESSION   BF254321
VERSION     BF254321.2  GI:13117068
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 527)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D.,
            Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W.,
            Fenton,R.D., Oates,R. and Main,D.
  TITLE     Development of a genetically and physically anchored EST resource
            for barley genomics: Morex unstressed seedling root cDNA library
  JOURNAL   Unpublished (2001)
COMMENT     On Nov 16, 2000 this sequence version replaced gi:11183426.
            Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total hq bases = 407
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence stop: 522.
FEATURES             Location/Qualifiers
     source          1. .527
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Morex"
                     /sub_species="vulgare"
                     /db_xref="taxon:112509"
```

/clone="HVSMEf0003K07f"
/tissue_type="Seedling root"
/lab_host="TJC121"
/clone_lib="Hordeum vulgare seedling root EST library HVcDNA0007 (Etiolated and unstressed)"
/note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1; Seeds were surface sterilized then germinated under axenic conditions in the dark at room temperature on filter paper with water, nystatin and cefotaxime in covered crystallization dishes. Five-day old seedling roots were then harvested, total RNA was prepared, poly(A) RNA was purified, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids. These steps were performed in the TJ Close laboratory at the University of California, Riverside (Choi, Close, Fenton). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

ORIGIN

Alignment Scores:

Pred. No.:	1.1	Length:	527
Score:	76.00	Matches:	11
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	90.48%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-15 (1-12) x BF254321 (1-527)

```
Qy          2 AsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||
Db        253 AACCACAACCACAACCACAACCACAACCACAAC 285
```

RESULT 175
BU494816
LOCUS BU494816 562 bp mRNA linear EST 11-SEP-2002
DEFINITION PfESToab68b03.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium falciparum 3D7 cDNA 5', mRNA sequence.
ACCESSION BU494816
VERSION BU494816.1 GI:22791010
KEYWORDS EST.
SOURCE Plasmodium falciparum 3D7
ORGANISM Plasmodium falciparum 3D7
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 562)
AUTHORS Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D., Marra,M., Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Jentes,E., Ronko,I., Tsagareishvili,R., Belaygorod,L., Franklin,C., Carr,L., Grow,A., Maguire,L., Richey,J., Wadkins,J., Kennedy,S., Levinso,D., Waterston,R., Wilson,R. and Sibley,D.
TITLE WashU Plasmodium EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: L. David Sibley
WashU Plasmodium EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library was constructed by Debopam Chakrabarti DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: L. David Sibley (sibley@borcim.wustl.edu), Washington University
Seq primer: -40UP from Gibco
High quality sequence stop: 424.
FEATURES Location/Qualifiers
source 1. .562
/organism="Plasmodium falciparum 3D7"
/mol_type="mRNA"
/db_xref="taxon:36329"
/lab_host="DH10B (GeneHog, Invitrogen, Inc.)"
/clone_lib="Plasmodium falciparum 3D7 asexual cDNA"
/note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2: XhoI; Library was constructed by Debopam Chakrabarti. Total RNA samples were isolated from mixed stage saponin(0.1%)-lysed P. falciparum 3D7 infected erythrocytes by the acidic guanidinium-phenol chloroform method. The poly A+ RNA was isolated by the polyAT-Tract mRNA isolation system (Promega, WI) using streptavidin Magnesphere particles. Directional cDNA libraries were constructed by oligo d(T) priming of poly(A)+ RNA (5mg) into EcoRI and XhoI sites of l ZapII vector using the Zap cDNA synthesis kit (Stragene, CA). The average size of the cDNA inserts in the library was between 1.0 and 1.5kb. Clones were mass excised using the ExAssist helper phage (Stratagene), the phagemids were precitptated with PEG 8000 and extracted with phenol/chloroform. Phagemid DNA was electroporated into DH10B cells."
ORIGIN

Alignment Scores:

Pred. No.:	1.17	Length:	562
Score:	76.00	Matches:	11
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	90.48%	Indels:	0
DB:	5	Gaps:	0

US-09-858-332G-15 (1-12) x BU494816 (1-562)

```
Qy          2 AsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||
Db         73 AATCATAATCATAATCATAATCATAATCATAAT 105
```

RESULT 176
BU997619
LOCUS BU997619 566 bp mRNA linear EST 23-OCT-2002
DEFINITION HI08I06r HI Hordeum vulgare subsp. vulgare cDNA clone HI08I06 5-PRIME, mRNA sequence.
ACCESSION BU997619
VERSION BU997619.1 GI:24274602
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 566)
AUTHORS Zhang,H., Weschke,W., Michalek,W., Stein,N. and Graner,A.
TITLE EST sequencing and analysis in barley (2002)
JOURNAL Unpublished (2002)
COMMENT Contact: Stein Nils
Molecular Markers Group, Department Genbank
Institute of Plant Genetics and Crop Plant Research (IPK)
Correnssstr. 3, 06466, Gatersleben, Germany
Tel: 039482-5522
Fax: 039482-5595
Email: stein@ipk-gatersleben.de
Insert Length: 566 Std Error: 0.00
Plate: 8 row: I column: 6
Seq primer: M13rev.
FEATURES Location/Qualifiers

```
     source          1. .566
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Barke"
                     /sub_species="vulgare"
                     /db_xref="GABI:251483"
                     /db_xref="taxon:112509"
                     /clone="HI08I06"
                     /tissue_type="female inflorescences"
                     /dev_stage="female inflorescences (approx. 3 mm in size)"
                     /lab_host="XL10-Gold"
                     /clone_lib="HI"
                     /note="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of
                     cDNA); Site_2: XhoI (3'-end of cDNA); Due to a cloning
                     artefact caused by the kit, in most cases the EcoRI site
                     is NOT present, as well as the EcoRIadapter used for
                     cloning. To excise the insert, restriction sites upstream
                     EcoRI should be used (e.g. BamHI, SalI,PstI). NOTE: Also
                     due to the cloning system used Blue/white selection for
                     recombinats is not 100% reliable."
ORIGIN

Alignment Scores:
Pred. No.:                1.18          Length:         566
Score:                    76.00         Matches:        11
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              90.48%        Indels:         0
DB:                       5             Gaps:           0

US-09-858-332G-15 (1-12) x BU997619 (1-566)

Qy          2 AsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||
Db        112 AACCACAACCACAACCACAACCACAACCACAAC 144

RESULT 177
CF638492
LOCUS       CF638492                 587 bp    mRNA    linear   EST 02-OCT-2003
DEFINITION  D03_C02 Filamentous Forced Diploid Ustilago maydis cDNA 3', mRNA
            sequence.
ACCESSION   CF638492
VERSION     CF638492.1  GI:37402226
KEYWORDS    EST.
SOURCE      Ustilago maydis
  ORGANISM  Ustilago maydis
            Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
            Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
REFERENCE   1  (bases 1 to 587)
  AUTHORS   Nugent,K.G., Choffe,K. and Saville,B.J.
  TITLE     Gene expression during Ustilago maydis diploid filamentous growth:
            EST library creation and analyses
  JOURNAL   Fungal Genet. Biol. 41 (3), 349-360 (2004)
   PUBMED   14761795
COMMENT     Contact: Barry J. Saville
            Saville Lab
            University of Toronto
            3359 Mississauga Road North, Mississauga, ON, L5L 1C6, Canada
            Tel: 905 569 4702
            Fax: 905 828 3792
            Email: bsaville@utm.utoronto.ca
            Plate: UTM-UM-D126/7-003-UTM  row: 02  column: C
            Seq primer: T7 Reverse (5' GAGTAATACGACTCACTATAGGG 3')
            High quality sequence stop: 587.
FEATURES             Location/Qualifiers
     source          1. .587
                     /organism="Ustilago maydis"
                     /mol_type="mRNA"
                     /strain="FBD12"
                     /db_xref="taxon:5270"
                     /cell_type="Mycelia"
                     /dev_stage="Filamentous diploid"
                     /clone_lib="Filamentous Forced Diploid"
                     /note="Vector: pSport; mRNA was extracted from diploid
                     mycelia. A cDNA library was constructed and
                     unidirectionally cloned into pSPORT plasmid, with the use
                     of the Superscript II cDNA Library Construction Kit."
ORIGIN

Alignment Scores:
Pred. No.:                1.23          Length:         587
Score:                    76.00         Matches:        11
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              90.48%        Indels:         0
DB:                       6             Gaps:           0

US-09-858-332G-15 (1-12) x CF638492 (1-587)

Qy          2 AsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||
Db        549 AACCACAACCACAACCACAACCACAACCACAAC 581

RESULT 178
CA855800
LOCUS       CA855800                 597 bp    mRNA    linear   EST 17-DEC-2002
DEFINITION  PfESToac38d05.y1 Plasmodium falciparum 3D7 gametocyte cDNA library
            Plasmodium falciparum 3D7 cDNA 5', mRNA sequence.
ACCESSION   CA855800
VERSION     CA855800.1  GI:27157911
KEYWORDS    EST.
SOURCE      Plasmodium falciparum 3D7
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 597)
  AUTHORS   Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D.,
            Marra,M., Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B.,
            Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Jentes,E., Ronko,I.,
            Tsagareishvili,R., Belaygorod,L., Franklin,C., Carr,L., Grow,A.,
            Maguire,L., Richey,J., Wadkins,J., Kennedy,S., Levinso,D.,
            Waterston,R., Wilson,R. and Sibley,D.
  TITLE     WashU Plasmodium EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: L. David Sibley
            WashU Plasmodium EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library was constructed by R. Haywood. DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: L. David Sibley
            (sibley@borcim.wustl.edu), Washington University
            Seq primer: -40UP from Gibco
            High quality sequence stop: 430.
FEATURES             Location/Qualifiers
     source          1. .597
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="mRNA"
                     /db_xref="taxon:36329"
                     /dev_stage="gametocyte (stage III-V)"
                     /lab_host="DH10B (GeneHog, Invitrogen, Inc.)"
                     /clone_lib="Plasmodium falciparum 3D7 gametocyte cDNA
                     library"
                     /note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2:
                     XhoI; The library was constructed by R Haywood. cDNAs were
                     synthesized  from ganetocyte poly(A)+ RNA by oligo d(T)
                     priming, size-selected  and directionally cloned into the
                     EcoRi (5' end) to XhoI (3' end)  sites of the Uni-ZAP XR
                     lambda vector (Stratagene). The primary  library was mass
                     excised as phagemid using the ExAssist helper  phage
                     (Stratagene). Clones were mass excised using the ExAssist
                     helper phage (Stratagene), the phagemids were precitptated
                     with  PEG 8000 and extracted with phenol/chloroform.
                     Phagemid DNA was  electroporated into DH10B cells. Clone
```

```
                    Availability: David Sibley,  Washington University."
ORIGIN

Alignment Scores:
Pred. No.:              1.25          Length:        597
Score:                  76.00         Matches:       11
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            90.48%        Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-15 (1-12) x CA855800 (1-597)

Qy          2 AsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||
Db        507 AATCATAATCATAATCATAATCATAATCATAAT 539

RESULT 179
CA854726
LOCUS       CA854726                600 bp    mRNA    linear   EST 17-DEC-2002
DEFINITION  PfESToac12f12.y1 Plasmodium falciparum 3D7 gametocyte cDNA library
            Plasmodium falciparum 3D7 cDNA 5', mRNA sequence.
ACCESSION   CA854726
VERSION     CA854726.1  GI:27156087
KEYWORDS    EST.
SOURCE      Plasmodium falciparum 3D7
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D.,
            Marra,M., Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B.,
            Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Jentes,E., Ronko,I.,
            Tsagareishvili,R., Belaygorod,L., Franklin,C., Carr,L., Grow,A.,
            Maguire,L., Richey,J., Wadkins,J., Kennedy,S., Levinso,D.,
            Waterston,R., Wilson,R. and Sibley,D.
  TITLE     WashU Plasmodium EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: L. David Sibley
            WashU Plasmodium EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library was constructed by R. Haywood. DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: L. David Sibley
            (sibley@borcim.wustl.edu), Washington University
            Seq primer: -40UP from Gibco
            High quality sequence stop: 432.
FEATURES             Location/Qualifiers
     source          1. .600
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="mRNA"
                     /db_xref="taxon:36329"
                     /dev_stage="gametocyte (stage III-V)"
                     /lab_host="DH10B (GeneHog, Invitrogen, Inc.)"
                     /clone_lib="Plasmodium falciparum 3D7 gametocyte cDNA
                     library"
                     /note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2:
                     XhoI; The library was constructed by R Haywood. cDNAs were
                     synthesized  from ganetocyte poly(A)+ RNA by oligo d(T)
                     priming, size-selected  and directionally cloned into the
                     EcoRi (5' end) to XhoI (3' end)  sites of the Uni-ZAP XR
                     lambda vector (Stratagene). The primary  library was mass
                     excised as phagemid using the ExAssist helper  phage
                     (Stratagene). Clones were mass excised using the ExAssist
                     helper phage (Stratagene), the phagemids were precitptated
                     with  PEG 8000 and extracted with phenol/chloroform.
                     Phagemid DNA was  electroporated into DH10B cells. Clone
                     Availability: David Sibley,  Washington University."
ORIGIN

Alignment Scores:
Pred. No.:              1.25          Length:        600
Score:                  76.00         Matches:       11
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            90.48%        Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-15 (1-12) x CA854726 (1-600)

Qy          2 AsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||
Db        507 AATCATAATCATAATCATAATCATAATCATAAT 539

RESULT 180
CA837970
LOCUS       CA837970                612 bp    mRNA    linear   EST 12-DEC-2002
DEFINITION  MCT011D06_165871 Ice plant Lambda Uni-Zap XR expression library, 5
            days 0.5 M NaCl treatment, Crassulacean acid metabolism, phase IV
            (5:30 PM). Mesembryanthemum crystallinum cDNA clone MCT011D06 5,
            mRNA sequence.
ACCESSION   CA837970
VERSION     CA837970.1  GI:26565735
KEYWORDS    EST.
SOURCE      Mesembryanthemum crystallinum (common iceplant)
  ORGANISM  Mesembryanthemum crystallinum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            Caryophyllales; Aizoaceae; Mesembryanthemum.
REFERENCE   1  (bases 1 to 612)
  AUTHORS   Cushman, J.C.
  TITLE     An expressed sequence tag database for the common ice plant,
            Mesembryanthemum crystallinum
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Cushman JC
            Department of Biochemistry
            University of Nevada
            MS200, Reno, NV 89557-0014, USA
            Tel: 775-784-1918
            Fax: 775-784-1650
            Email: jcushman@unr.edu
            PCR PRimers
            FORWARD: T3 20mer
            BACKWARD: T7 21mer
            Plate: 011  row: D  column: 06
            Seq primer: T3 20mer
            High quality sequence stop: 612.
FEATURES             Location/Qualifiers
     source          1. .612
                     /organism="Mesembryanthemum crystallinum"
                     /mol_type="mRNA"
                     /db_xref="taxon:3544"
                     /clone="MCT011D06"
                     /tissue_type="leaf"
                     /dev_stage="five-week-old"
                     /clone_lib="Ice plant Lambda Uni-Zap XR expression
                     library, 5 days 0.5 M NaCl treatment, Crassulacean acid
                     metabolism, phase IV (5:30 PM)."
                     /note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Library construction was performed
                     according to Stratagene's recommended protocol for the
                     Lambda UniZapXR vector and cDNA synthesis kit."
ORIGIN

Alignment Scores:
Pred. No.:              1.28          Length:        612
Score:                  76.00         Matches:       11
Percent Similarity:     91.67%        Conservative:  0
Best Local Similarity:  91.67%        Mismatches:    1
Query Match:            90.48%        Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-15 (1-12) x CA837970 (1-612)
```

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||   |||||||||
Db        153 CACAACCACAACCACAACCACAACCGCAACCACAAC 188

RESULT 181
CA855323
LOCUS       CA855323                671 bp    mRNA    linear   EST 17-DEC-2002
DEFINITION  PfESToac36c03.y1 Plasmodium falciparum 3D7 gametocyte cDNA library
            Plasmodium falciparum 3D7 cDNA 5', mRNA sequence.
ACCESSION   CA855323
VERSION     CA855323.1  GI:27157096
KEYWORDS    EST.
SOURCE      Plasmodium falciparum 3D7
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 671)
  AUTHORS   Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D.,
            Marra,M., Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B.,
            Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Jentes,E., Ronko,I.,
            Tsagareishvili,R., Belaygorod,L., Franklin,C., Carr,L., Grow,A.,
            Maguire,L., Richey,J., Wadkins,J., Kennedy,S., Levinso,D.,
            Waterston,R., Wilson,R. and Sibley,D.
  TITLE     WashU Plasmodium EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: L. David Sibley
            WashU Plasmodium EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library was constructed by R. Haywood. DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: L. David Sibley
            (sibley@borcim.wustl.edu), Washington University
            Seq primer: -40UP from Gibco
            High quality sequence stop: 438.
FEATURES             Location/Qualifiers
     source          1. .671
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="mRNA"
                     /db_xref="taxon:36329"
                     /dev_stage="gametocyte (stage III-V)"
                     /lab_host="DH10B (GeneHog, Invitrogen, Inc.)"
                     /clone_lib="Plasmodium falciparum 3D7 gametocyte cDNA
                     library"
                     /note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2:
                     XhoI; The library was constructed by R Haywood. cDNAs were
                     synthesized  from ganetocyte poly(A)+ RNA by oligo d(T)
                     priming, size-selected  and directionally cloned into the
                     EcoRi (5' end) to XhoI (3' end)  sites of the Uni-ZAP XR
                     lambda vector (Stratagene). The primary  library was mass
                     excised as phagemid using the ExAssist helper  phage
                     (Stratagene). Clones were mass excised using the ExAssist
                     helper phage (Stratagene), the phagemids were precitptated
                     with  PEG 8000 and extracted with phenol/chloroform.
                     Phagemid DNA was  electroporated into DH10B cells. Clone
                     Availability: David Sibley,  Washington University."
ORIGIN

Alignment Scores:
Pred. No.:                1.41          Length:          671
Score:                    76.00         Matches:         11
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              90.48%        Indels:          0
DB:                       6             Gaps:            0

US-09-858-332G-15 (1-12) x CA855323 (1-671)

Qy          2 AsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||
Db        497 AATCATAATCATAATCATAATCATAATCATAAT 529

RESULT 182
BE060493
LOCUS       BE060493                697 bp    mRNA    linear   EST 22-OCT-2001
DEFINITION  HVSMEg0012H07f Hordeum vulgare pre-anthesis spike EST library
            HVcDNA0008 (white to yellow anther) Hordeum vulgare subsp. vulgare
            cDNA clone HVSMEg0012H07f, mRNA sequence.
ACCESSION   BE060493
VERSION     BE060493.2  GI:13154324
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 697)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D.,
            Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W.,
            Fenton,R.D., Close,S.J., Oates,R. and Main,D.
  TITLE     Development of a genetically and physically anchored EST resource
            for barley genomics: Morex pre-anthesis spike cDNA library
  JOURNAL   Unpublished (2001)
COMMENT     On Jun 9, 2000 this sequence version replaced gi:8405143.
            Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total hq bases = 204
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence start: 2
            High quality sequence stop: 542.
FEATURES             Location/Qualifiers
     source          1. .697
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Morex"
                     /sub_species="vulgare"
                     /db_xref="taxon:112509"
                     /clone="HVSMEg0012H07f"
                     /tissue_type="pre-anthesis spike"
                     /lab_host="SOLR"
                     /clone_lib="Hordeum vulgare pre-anthesis spike EST library
                     HVcDNA0008 (white to yellow anther)"
                     /note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1;
                     Plants were grown in the greenhouse at the University of
                     California, Riverside (Fenton, SJ Close, TJ Close). Whole
                     spike with awns trimmed were collected at white, green and
                     yellow anther stages (Fenton). Total RNA was prepared from
                     each pool, equal quantities of all three RNA pools were
                     combined, poly(A) RNA was purified from the mixture, one
                     primary unamplified cDNA library was made, and 1 million
                     pfu were in vivo excised to give pBluescript SK(-) cDNA
                     phagemids. These steps were performed in the TJ Close lab
                     (Choi) at the University of California, Riverside.
                     Phagemids were plated and picked at the Clemson University
                     Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins
                     and Wing) Plasmid DNA preparations, DNA sequencing and
                     sequence analysis were performed at CUGI (Wing, Yu,
                     Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence
                     has been trimmed to remove vector sequence and contains a
                     minimum of 100 bases of phred value 20 or above. For more
                     details on library preparation and sequence analysis see
                     http://www.genome.clemson.edu/projects/barley. To order
                     this clone see http://www.genome.clemson.edu/orders Also
                     see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
                     Genetically and physically anchored EST resources for
                     barley genomics. Barley Genetics Newsletter 31:29-30.
                     (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
ORIGIN
```

```
Alignment Scores:
Pred. No.:                1.47          Length:         697
Score:                    76.00         Matches:        11
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              90.48%        Indels:         0
DB:                       2             Gaps:           0

US-09-858-332G-15 (1-12) x BE060493 (1-697)

Qy          2 AsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||
Db        256 AACCACAACCACAACCACAACCACAACCACAAC 288

RESULT 183
BU561234
LOCUS       BU561234                727 bp    mRNA    linear   EST 16-SEP-2002
DEFINITION  AGENCOURT_10279346 NIH_MGC_82 Homo sapiens cDNA clone IMAGE:6592195
            5', mRNA sequence.
ACCESSION   BU561234
VERSION     BU561234.1  GI:22911530
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 727)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2814  row: p  column: 19
            High quality sequence stop: 550.
FEATURES             Location/Qualifiers
     source          1. .727
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6592195"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_82"
                     /note="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).   Average insert size
                     1.35 kb (range 0.9-4.0 kb).  14/15 colonies contained
                     inserts by PCR. This library was enriched for full-length
                     clones and was constructed by Clontech  Laboratories (Palo
                     Alto, CA)."
ORIGIN

Alignment Scores:
Pred. No.:                1.53          Length:         727
Score:                    76.00         Matches:        11
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              90.48%        Indels:         0
DB:                       5             Gaps:           0

US-09-858-332G-15 (1-12) x BU561234 (1-727)

Qy          2 AsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||
Db        289 AATCATAATCATAATCATAATCATAATCATAAT 321

RESULT 184
CF638604
LOCUS       CF638604                745 bp    mRNA    linear   EST 02-OCT-2003
DEFINITION  D04_F10 Filamentous Forced Diploid Ustilago maydis cDNA 3', mRNA
            sequence.
ACCESSION   CF638604
VERSION     CF638604.1  GI:37402446
KEYWORDS    EST.
SOURCE      Ustilago maydis
  ORGANISM  Ustilago maydis
            Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
            Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
REFERENCE   1  (bases 1 to 745)
  AUTHORS   Nugent,K.G., Choffe,K. and Saville,B.J.
  TITLE     Gene expression during Ustilago maydis diploid filamentous growth:
            EST library creation and analyses
  JOURNAL   Fungal Genet. Biol. 41 (3), 349-360 (2004)
   PUBMED   14761795
COMMENT     Contact: Barry J. Saville
            Saville Lab
            University of Toronto
            3359 Mississauga Road North, Mississauga, ON, L5L 1C6, Canada
            Tel: 905 569 4702
            Fax: 905 828 3792
            Email: bsaville@utm.utoronto.ca
            Plate: UTM-UM-D126/7-004-UTM  row: 10  column: F
            Seq primer: T7 Reverse (5' GAGTAATACGACTCACTATAGGG 3')
            High quality sequence stop: 745.
FEATURES             Location/Qualifiers
     source          1. .745
                     /organism="Ustilago maydis"
                     /mol_type="mRNA"
                     /strain="FBD12"
                     /db_xref="taxon:5270"
                     /cell_type="Mycelia"
                     /dev_stage="Filamentous diploid"
                     /clone_lib="Filamentous Forced Diploid"
                     /note="Vector: pSport; mRNA was extracted from diploid
                     mycelia. A cDNA library was constructed and
                     unidirectionally cloned into pSPORT plasmid, with the use
                     of the Superscript II cDNA Library Construction Kit."
ORIGIN

Alignment Scores:
Pred. No.:                1.57          Length:         745
Score:                    76.00         Matches:        11
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              90.48%        Indels:         0
DB:                       6             Gaps:           0

US-09-858-332G-15 (1-12) x CF638604 (1-745)

Qy          2 AsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||
Db        549 AACCACAACCACAACCACAACCACAACCACAAC 581

RESULT 185
BF235379
LOCUS       BF235379                748 bp    mRNA    linear   EST 14-NOV-2000
DEFINITION  602026773F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:4161714 5',
            mRNA sequence.
ACCESSION   BF235379
VERSION     BF235379.1  GI:11147878
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
```

```
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 748)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9443  row: b  column: 19
            High quality sequence stop: 742.
FEATURES              Location/Qualifiers
     source           1. .748
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /strain="FVB/N"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:4161714"
                      /lab_host="DH10B (T1 phage-resistant)"
                      /clone_lib="NCI_CGAP_Li9"
                      /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                      Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                      Average insert size 1.9 kb. Constructed by Life
                      Technologies. Note: this is a NCI_CGAP Library."
ORIGIN

Alignment Scores:
Pred. No.:                1.58          Length:          748
Score:                    76.00         Matches:         11
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              90.48%        Indels:          0
DB:                       2             Gaps:            0

US-09-858-332G-15 (1-12) x BF235379 (1-748)

Qy          2 AsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||
Db          3 AACCACAACCACAACCACAACCACAACCACAAC 35

RESULT 186
BU275786
LOCUS       BU275786                322 bp    mRNA    linear   EST 03-JAN-2003
DEFINITION  Cr_Emb_07E01_TEXF1 Convoluta roscoffensis embryos from Eva Jiminez
            Symsagittifera roscofensis cDNA clone Cr_Emb_07E01 5', mRNA
            sequence.
ACCESSION   BU275786
VERSION     BU275786.1  GI:27476556
KEYWORDS    EST.
SOURCE      Symsagittifera roscofensis
  ORGANISM  Symsagittifera roscofensis
            Eukaryota; Metazoa; Platyhelminthes; Turbellaria; Acoelomorpha;
            Acoela; Sagittiferidae; Symsagittifera.
REFERENCE   1  (bases 1 to 322)
  AUTHORS   Blaxter,M.L., Jiminez,E., Rosie,A., Aboobaker,A. and Salo,E.
  TITLE     A survey of genes expressed in the acoel flatworm Convoluta
            roscoffensis
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Blaxter ML
            Institute of Cell, Animal and Population Biology
            University of Edinburgh
            Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
            3JT, UK.
            Tel: +44 131 650 6760
            Fax: +44 131 670 5450
            Email: mark.blaxter@ed.ac.uk
            The library was prepared by Eva Jimenez, University of Barcelona,
            Spain The sequencing was performed by Eva Jimenez, Aziz Aboobaker
            and Ailie Rosie Edinburgh Phylogenomics Programme, ICAPB,
            Edinburgh, UK.The sequence contained a PolyA tail (trimmed)
            PCR PRimers
            FORWARD: T7PL
            BACKWARD: triple 5 Exseq
            Plate: 07  row: E  column: 01
            Seq primer: TEXF1
            High quality sequence stop: 322.
FEATURES              Location/Qualifiers
     source           1. .322
                      /organism="Symsagittifera roscofensis"
                      /mol_type="mRNA"
                      /db_xref="taxon:84072"
                      /clone="Cr_Emb_07E01"
                      /sex="mixed"
                      /dev_stage="embryo"
                      /clone_lib="Convoluta roscoffensis embryos from Eva
                      Jiminez"
                      /note="Vector: lambda triplex; Site_1: EcoRI; Site_2:
                      HinDIII; The library was prepared by Eva Jimenez,
                      University of Barcelona, Spain The sequencing was
                      performed by Eva Jimenez,  Aziz Aboobaker and Ailie Rosie,
                      Edinburgh Phylogenomics Programme,  ICAPB, Edinburgh, UK."
ORIGIN

Alignment Scores:
Pred. No.:                0.88          Length:          322
Score:                    75.00         Matches:         11
Percent Similarity:       91.67%        Conservative:    0
Best Local Similarity:    91.67%        Mismatches:      1
Query Match:              89.29%        Indels:          0
DB:                       5             Gaps:            0

US-09-858-332G-15 (1-12) x BU275786 (1-322)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||||||||||||||||||||||   |||||||||
Db        265 CATAACCATAACCATAACCATAATAGTAATCATAAC 300

RESULT 187
CNS026FK/c
LOCUS       CNS026FK                487 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            240J20 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL183305
VERSION     AL183305.1  GI:7821409
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
   PUBMED   10899143
REFERENCE   3  (bases 1 to 487)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .487
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="240J20"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG240DE10LP1
                     end : T7"
ORIGIN

Alignment Scores:
Pred. No.:                1.36          Length:        487
Score:                    75.00         Matches:       11
Percent Similarity:       91.67%        Conservative:  0
Best Local Similarity:    91.67%        Mismatches:    1
Query Match:              89.29%        Indels:        0
DB:                       10            Gaps:          0

US-09-858-332G-15 (1-12) x CNS026FK (1-487)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||   |||
Db        136 CACAACCACAACCACAACCACAACCACAACCRGAAC 101

RESULT 188
CW461432
LOCUS       CW461432                655 bp    DNA     linear   GSS 02-NOV-2004
DEFINITION  fsbb001f209b13f0 Sorghum methylation filtered library (LibID: 104)
            Sorghum bicolor genomic clone fsbb001f209b13, genomic survey
            sequence.
ACCESSION   CW461432
VERSION     CW461432.1  GI:55209393
KEYWORDS    GSS.
SOURCE      Sorghum bicolor (sorghum)
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 655)
  AUTHORS   Bedell,J.A., Budiman,M.A., Nunberg,A., Citek,R.W., Robbins,D.,
            Jones,J., Flick,E., Rohlfing,T., Fries,J., Bradford,K.,
            McMenamy,J., Smith,M., Holeman,H., Roe,B.A, Wiley,G., Korf,I.F.,
            Rabinowicz,P.D., Lakey,N., McCombie,W.R., Jeddeloh,J.A. and
            Martienssen,R.A.
  TITLE     Sorghum genome sequencing by methylation filtration
  JOURNAL   PLoS Biol. 3 (1), e13 (2005)
   PUBMED   15660154
COMMENT     Contact: Bedell JA
            Orion Genomics, LLC
            4041 Forest Park Ave, St. Louis, MO 63108, USA
            Tel: 314 615 6979
            Fax: 314 615 5975
            Email: jbedell@oriongenomics.com
            Plate: fsbb001f209  row: b  column: 13
            Seq primer: f Forward
            Class: methylation filtered
            High quality sequence stop: 655.
FEATURES             Location/Qualifiers
     source          1. .655
                     /organism="Sorghum bicolor"
                     /mol_type="genomic DNA"
                     /cultivar="ATx623"
                     /db_xref="taxon:4558"
                     /clone="fsbb001f209b13"
                     /clone_lib="Sorghum methylation filtered library (LibID:
                     104)"
                     /note="Organ: leaf; Vector: pBCSK(-); Site_1: HincII; DNA
                     prepared from purified nuclei was randomly sheared,
                     end-repaired, size fractionated to enrich for the 0.5 to 5
                     kb fraction, ligated into HincII-digested pBCSK(-) vector
                     and electroporated into E. coli cells. This is a
                     methylation filtered library."
ORIGIN

Alignment Scores:
Pred. No.:                1.85          Length:        655
Score:                    75.00         Matches:       11
Percent Similarity:       91.67%        Conservative:  0
Best Local Similarity:    91.67%        Mismatches:    1
Query Match:              89.29%        Indels:        0
DB:                       10            Gaps:          0

US-09-858-332G-15 (1-12) x CW461432 (1-655)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||   |||
Db        139 CACAATCACAATCACAATCACAATCACAATGACAAT 174

RESULT 189
CW233590
LOCUS       CW233590                662 bp    DNA     linear   GSS 30-OCT-2004
DEFINITION  104_687_11213527_148_37378_032 Sorghum methylation filtered library
            (LibID: 104) Sorghum bicolor genomic clone 11213527, genomic survey
            sequence.
ACCESSION   CW233590
VERSION     CW233590.1  GI:54942417
KEYWORDS    GSS.
SOURCE      Sorghum bicolor (sorghum)
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 662)
  AUTHORS   Bedell,J.A., Budiman,M.A., Nunberg,A., Citek,R.W., Robbins,D.,
            Jones,J., Flick,E., Rohlfing,T., Fries,J., Bradford,K.,
            McMenamy,J., Smith,M., Holeman,H., Roe,B.A, Wiley,G., Korf,I.F.,
            Rabinowicz,P.D., Lakey,N., McCombie,W.R., Jeddeloh,J.A. and
            Martienssen,R.A.
  TITLE     Sorghum genome sequencing by methylation filtration
  JOURNAL   PLoS Biol. 3 (1), e13 (2005)
   PUBMED   15660154
COMMENT     Contact: Bedell JA
            Orion Genomics, LLC
            4041 Forest Park Ave, St. Louis, MO 63108, USA
            Tel: 314 615 6979
            Fax: 314 615 5975
            Email: jbedell@oriongenomics.com
            Plate: 687  row: a  column: 07
            Seq primer: SWfor Forward
            Class: methylation filtered
            High quality sequence stop: 662.
FEATURES             Location/Qualifiers
     source          1. .662
                     /organism="Sorghum bicolor"
                     /mol_type="genomic DNA"
                     /cultivar="ATx623"
                     /db_xref="taxon:4558"
                     /clone="11213527"
                     /clone_lib="Sorghum methylation filtered library (LibID:
                     104)"
                     /note="Organ: leaf; Vector: pBCSK(-); Site_1: HincII; DNA
                     prepared from purified nuclei was randomly sheared,
                     end-repaired, size fractionated to enrich for the 0.5 to 5
                     kb fraction, ligated into HincII-digested pBCSK(-) vector
                     and electroporated into E. coli cells. This is a
                     methylation filtered library."
ORIGIN
```

```
Alignment Scores:
Pred. No.:                1.87          Length:          662
Score:                    75.00         Matches:         11
Percent Similarity:       91.67%        Conservative:    0
Best Local Similarity:    91.67%        Mismatches:      1
Query Match:              89.29%        Indels:          0
DB:                       10            Gaps:            0

US-09-858-332G-15 (1-12) x CW233590 (1-662)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||   |||
Db        265 CACAATCACAATCACAATCACAATCACAATGACAAT 300

RESULT 190
CL405652/c
LOCUS       CL405652                  784 bp    DNA     linear   GSS 05-MAR-2004
DEFINITION  ZMMBBb0397J19r ZMMBBb (HindIII) Zea mays genomic clone
            ZMMBBb0397J19 3', genomic survey sequence.
ACCESSION   CL405652
VERSION     CL405652.1  GI:45216871
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 784)
  AUTHORS   Bharti,A.K., Young,S., Kavchok,S., Keizer,G., Bronzino,A.C.,
            Zohovetz,V., Fuks,G., Yu,Y., Wing,R. and Messing,J.
  TITLE     Sequencing of the maize genome at PGIR (2003c)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Bharti,A.K.
            Dr.Joachim Messing's lab
            The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers
            University
            190 Frelinghuysen Road, Piscataway, NJ 08854, USA
            Tel: 732 445 3801
            Fax: 732 445 5735
            Email: bharti@waksman.rutgers.edu
            Seq primer: SP6
            Class: BAC ends
            High quality sequence start: 416.
FEATURES             Location/Qualifiers
     source          1. .784
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBBb0397J19"
                     /lab_host="E. coli DH10B"
                     /clone_lib="ZMMBBb (HindIII)"
                     /note="Vector: pCUGI; Site_1: HindIII; Site_2: HindIII"
ORIGIN

Alignment Scores:
Pred. No.:                2.23          Length:          784
Score:                    75.00         Matches:         11
Percent Similarity:       91.67%        Conservative:    0
Best Local Similarity:    91.67%        Mismatches:      1
Query Match:              89.29%        Indels:          0
DB:                       10            Gaps:            0

US-09-858-332G-15 (1-12) x CL405652 (1-784)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||   |||
Db        296 CATAATCATAATCATAATCATAATCATAATGACAAC 261

RESULT 191
CL960835
LOCUS       CL960835                  291 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFCC037313 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL960835
VERSION     CL960835.1  GI:52376381
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

Alignment Scores:
Pred. No.:                1.06          Length:          291
Score:                    74.00         Matches:         10
Percent Similarity:       100.00%       Conservative:    2
Best Local Similarity:    83.33%        Mismatches:      0
Query Match:              88.10%        Indels:          0
DB:                       10            Gaps:            0

US-09-858-332G-15 (1-12) x CL960835 (1-291)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||
Db        145 CACCACCACGACCACAACCACAACCACAACCACAAC 180

RESULT 192
CF963928
LOCUS       CF963928                  368 bp    mRNA    linear   EST 15-JAN-2004
DEFINITION  13692rsiced_4143.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF963928
VERSION     CF963928.1  GI:38479839
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 368)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
  PUBMED    15685292
```

```
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 368
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .368
                      /organism="Oryza sativa (indica cultivar-group)"
                      /mol_type="mRNA"
                      /cultivar="LYP9"
                      /db_xref="taxon:39946"
                      /tissue_type="whole plant"
                      /dev_stage="trefoil"
                      /clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA
                      library"
ORIGIN

Alignment Scores:
Pred. No.:                1.36          Length:          368
Score:                    74.00         Matches:         10
Percent Similarity:       100.00%       Conservative:    2
Best Local Similarity:    83.33%        Mismatches:      0
Query Match:              88.10%        Indels:          0
DB:                       7             Gaps:            0

US-09-858-332G-15 (1-12) x CF963928 (1-368)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        206 CACCACCACGACCACAACCACAACCACAACCACAAC 241

RESULT 193
BF068492
LOCUS       BF068492                386 bp    mRNA    linear   EST 13-JUL-2004
DEFINITION  st82f08.y1 Gm-c1054 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1054-1024 5' similar to SW:IF4Z_TOBAC Q40468 EUKARYOTIC
            INITIATION FACTOR 4A-15 ;, mRNA sequence.
ACCESSION   BF068492
VERSION     BF068492.1  GI:10845345
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 386)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,
            Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
            Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
            Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
            Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            When it has been determined, an EST from the other end of this
            clone is listed in the 'Other ESTs on clone' field. This clone is
            available through: Biogenetic Services, 801 32nd Ave. Brookings, SD
            57006 USA (phone: 800 423 4163; email:
            info@biogeneticservices.com).
FEATURES              Location/Qualifiers
     source           1. .386
                      /organism="Glycine max"
                      /mol_type="mRNA"
                      /cultivar="Harosoy"
                      /db_xref="taxon:3847"
                      /clone="GENOME SYSTEMS CLONE ID: Gm-c1054-1024"
                      /tissue_type="Leaf, 3 week old, greenhouse grown"
                      /lab_host="DH10B"
                      /clone_lib="Gm-c1054"
                      /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                      XhoI; The Harosoy NIL was constructed and seed was
                      provided by Dr. J. Specht, University of Nebraska
                      (Shoemaker and Specht, 1995). The cDNA library was
                      constructed from mRNA isolated from leaf tissue at various
                      developmental stages of 3 week old greenhouse grown
                      plants. Complementary DNA was  synthesized from mRNA using
                      a primer consisting of a poly(dT) sequence  with a XhoI
                      restriction site and a 3' anchor. EcoRI adapters were
                      ligated  to the blunt-ended cDNA fragments followed by
                      XhoI digestion. The cDNA  fragments were directionally
                      cloned into the EcoRI-XhoI restriction site of  the
                      pBluescript vector. The ligated cDNA fragments were
                      transformed into  DH10B host cells (GibcoBRL). This
                      library was constructed in cooperation  with Dr. Paul
                      Keim's laboratory at Northern Arizona University."
ORIGIN

Alignment Scores:
Pred. No.:                1.43          Length:          386
Score:                    74.00         Matches:         10
Percent Similarity:       100.00%       Conservative:    2
Best Local Similarity:    83.33%        Mismatches:      0
Query Match:              88.10%        Indels:          0
DB:                       2             Gaps:            0

US-09-858-332G-15 (1-12) x BF068492 (1-386)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db         27 CACAGCCACAGCCACAACCACAACCACAACCACAAC 62

RESULT 194
CF196074
LOCUS       CF196074                395 bp    mRNA    linear   EST 31-JUL-2003
DEFINITION  Azbi2.3_E03 AZ-BI Oryza sativa (japonica cultivar-group) cDNA clone
            AZ-BI2.3E03 5', mRNA sequence.
ACCESSION   CF196074
VERSION     CF196074.1  GI:33380731
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 395)
  AUTHORS   Ventelon-Debout,M., Thi-Thu Huang,N., Wissocq,A., Berger,C.,
            Michele,L., Piegu,B., Cooke,R., Ghesquiere,A., Delseny,M. and
            Brugidou,C.
  TITLE     Metabolism and Photosynthesis gene expression is affected in
            response to rice yellow mottle virus infection in Oryza sativa
            indica and japonica culltivars
  JOURNAL   Mol. Genet. Genomics (2003) In press
COMMENT     Contact: Brugidou C
            IRD
            BP 64501, 34394 Montpellier cedex 5 France
            Tel: 33 4 67 41 67 39
            Fax: 33 4 67 41 61 81
            PCR PRimers
            FORWARD: 5'AATTAACCCTCACTAAAGGG3'
            BACKWARD: 5'CGGGATATCACTCAGCATAATG3'
            Plate: 2.3  row: E  column: 3.
FEATURES              Location/Qualifiers
     source           1. .395
```

/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Azucena"
/db_xref="taxon:39947"
/clone="AZ-BI2.3E03"
/tissue_type="leaves"
/dev_stage="3 weeks-old plants"
/lab_host="SOLR strain (Stratagene)"
/clone_lib="AZ-BI"
/note="Vector: pBluescript SK (+/-); Site_1: Eco RI; Site_2: Xho I; Sample name: AZ-BI; Plant growth place: greenhouse; Soil conditions: standard; Sowing date: 28/02/2000; Harvesting date: 18/03/2000; Stress date: 15/03/2000; Cultivar: azucena; Description: Az-wounded leaves cDNA library; experimental condition: wound;"
ORIGIN

Alignment Scores:

Pred. No.:	1.47	Length:	395
Score:	74.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	83.33%	Mismatches:	0
Query Match:	88.10%	Indels:	0
DB:	6	Gaps:	0

US-09-858-332G-15 (1-12) x CF196074 (1-395)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db         76 CACCACCACGACCACAACCACAACCACAACCACAAC 111
```

RESULT 195
CF195899
LOCUS CF195899 405 bp mRNA linear EST 31-JUL-2003
DEFINITION Azbi2.1_A12 AZ-BI Oryza sativa (japonica cultivar-group) cDNA clone AZ-BI2.1A12 5', mRNA sequence.
ACCESSION CF195899
VERSION CF195899.1 GI:33380556
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 405)
AUTHORS Ventelon-Debout,M., Thi-Thu Huang,N., Wissocq,A., Berger,C., Michele,L., Piegu,B., Cooke,R., Ghesquiere,A., Delseny,M. and Brugidou,C.
TITLE Metabolism and Photosynthesis gene expression is affected in response to rice yellow mottle virus infection in Oryza sativa indica and japonica culltivars
JOURNAL Mol. Genet. Genomics (2003) In press
COMMENT Contact: Brugidou C
IRD
BP 64501, 34394 Montpellier cedex 5 France
Tel: 33 4 67 41 67 39
Fax: 33 4 67 41 61 81
PCR PRimers
FORWARD: 5'AATTAACCCTCACTAAAGGG3'
BACKWARD: 5'CGGGATATCACTCAGCATAATG3'
Plate: 2.1 row: A column: 12.
FEATURES Location/Qualifiers
source 1. .405
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Azucena"
/db_xref="taxon:39947"
/clone="AZ-BI2.1A12"
/tissue_type="leaves"
/dev_stage="3 weeks-old plants"
/lab_host="SOLR strain (Stratagene)"
/clone_lib="AZ-BI"
/note="Vector: pBluescript SK (+/-); Site_1: Eco RI; Site_2: Xho I; Sample name: AZ-BI; Plant growth place: greenhouse; Soil conditions: standard; Sowing date: 28/02/2000; Harvesting date: 18/03/2000; Stress date: 15/03/2000; Cultivar: azucena; Description: Az-wounded leaves cDNA library; experimental condition: wound;"
ORIGIN

Alignment Scores:

Pred. No.:	1.5	Length:	405
Score:	74.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	83.33%	Mismatches:	0
Query Match:	88.10%	Indels:	0
DB:	6	Gaps:	0

US-09-858-332G-15 (1-12) x CF195899 (1-405)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db         41 CACCACCACGACCACAACCACAACCACAACCACAAC 76
```

RESULT 196
AQ655474/c
LOCUS AQ655474 420 bp DNA linear GSS 22-JUN-1999
DEFINITION Sheared DNA-25D5.TF Sheared DNA Trypanosoma brucei genomic clone Sheared DNA-25D5, genomic survey sequence.
ACCESSION AQ655474
VERSION AQ655474.1 GI:5148660
KEYWORDS GSS.
SOURCE Trypanosoma brucei
ORGANISM Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
REFERENCE 1 (bases 1 to 420)
AUTHORS El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C., Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J., Fraser,C. and Adams,M.
TITLE Determination of clone end sequences from Trypanosoma brucei GUTat 10.1 sheared DNA library
JOURNAL Unpublished (1999)
COMMENT Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared DNA library constructed at TIGR. Clones will be available for distribution through ATCC. Sheared DNA end sequences search page: http://www.tigr.org/tdb/mdb/tbdb/.
Seq primer: M13-Forward
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .420
/organism="Trypanosoma brucei"
/mol_type="genomic DNA"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone="Sheared DNA-25D5"
/clone_lib="Sheared DNA"
/note="Vector: pUC18; Site_1: SmaI; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (approx 2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barell, Oxford University Press, 1999)."
ORIGIN

```
Alignment Scores:
Pred. No.:                1.56          Length:         420
Score:                    74.00         Matches:        10
Percent Similarity:       100.00%       Conservative:   2
Best Local Similarity:    83.33%        Mismatches:     0
Query Match:              88.10%        Indels:         0
DB:                       9             Gaps:           0

US-09-858-332G-15 (1-12) x AQ655474 (1-420)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        353 CACCACCACCACCACAATCACAACCACAACCACAAC 318

RESULT 197
BI798061
LOCUS       BI798061                426 bp    mRNA    linear   EST 02-OCT-2001
DEFINITION  H096C06 Endosperm library from Oryza sativa (10 days after
            anthesis) Oryza sativa cDNA clone H096C06, mRNA sequence.
ACCESSION   BI798061
VERSION     BI798061.1  GI:15849785
KEYWORDS    EST.
SOURCE      Oryza sativa
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 426)
  AUTHORS   Dong,H.T., Li,D.B., Zhuang,X.F., Dai,C.G., Sun,L.X., Pei,Y.X.,
            Wu,H.F., Jiang,Y.X., Yu,F.C., Gao,Q.K. and Lou,Y.C.
  TITLE     A Gene Expression Screen in Oryza sativa
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Haitao Dong, Debao Li
            Bioinfomatics and Gene Network Research Group
            Zhejiang University
            Kaixuan Road 268#, Hangzhou, Zhejiang, P.R.China
            Tel: 0086-571-86892051
            Fax: 0086-571-86961525
            Email: webmaster@estarray.org, URL: http://www.estarray.org
            Seq primer: M13 forward primer.
FEATURES             Location/Qualifiers
     source          1. .426
                     /organism="Oryza sativa"
                     /mol_type="mRNA"
                     /db_xref="taxon:4530"
                     /clone="H096C06"
                     /tissue_type="Endosperm"
                     /dev_stage="10 days after anthesis"
                     /clone_lib="Endosperm library from Oryza sativa (10 days
                     after anthesis)"
                     /note="Vector: pSport2"
ORIGIN

Alignment Scores:
Pred. No.:                1.59          Length:         426
Score:                    74.00         Matches:        10
Percent Similarity:       100.00%       Conservative:   2
Best Local Similarity:    83.33%        Mismatches:     0
Query Match:              88.10%        Indels:         0
DB:                       3             Gaps:           0

US-09-858-332G-15 (1-12) x BI798061 (1-426)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db         37 CACCACCACGACCACAACCACAACCACAACCACAAC 72

RESULT 198
CF987483
LOCUS       CF987483                430 bp    mRNA    linear   EST 04-FEB-2005
DEFINITION  15498rsicee_11376.y1 Oryza sativa cv. LYP9 tillering whole plant
            cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF987483
VERSION     CF987483.1  GI:58589175
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 430)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 430
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .430
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="tillering"
                     /clone_lib="Oryza sativa cv. LYP9 tillering whole plant
                     cDNA library"
ORIGIN

Alignment Scores:
Pred. No.:                1.6           Length:         430
Score:                    74.00         Matches:        10
Percent Similarity:       100.00%       Conservative:   2
Best Local Similarity:    83.33%        Mismatches:     0
Query Match:              88.10%        Indels:         0
DB:                       7             Gaps:           0

US-09-858-332G-15 (1-12) x CF987483 (1-430)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        205 CACCACCACGACCACAACCACAACCACAACCACAAC 240

RESULT 199
CF965114
LOCUS       CF965114                437 bp    mRNA    linear   EST 15-JAN-2004
DEFINITION  9492rsiced_8168.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF965114
VERSION     CF965114.1  GI:38481025
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 437)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
```

Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J., Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q., Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z., Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P., Wu,S. and Liu,J.
TITLE The Genomes of Oryza sativa: A History of Duplications
JOURNAL PLoS Biol. 3 (2), e38 (2005)
PUBMED 15685292
COMMENT Contact: Yan Zhou
Bioinformatics Department
Hangzhou Genomics Institute
No.51 Zhijiang Road, Hangzhou 310008, China
Tel: 86-571-56805886
Fax: 86-571-56805884
Email: zhouyan@genomics.org.cn
Seq primer: M13 Forward
High quality sequence stop: 437
POLYA=No.

FEATURES Location/Qualifiers
source 1. .437
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/cultivar="LYP9"
/db_xref="taxon:39946"
/tissue_type="whole plant"
/dev_stage="trefoil"
/clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA library"

ORIGIN

Alignment Scores:

Pred. No.:	1.63	Length:	437
Score:	74.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	83.33%	Mismatches:	0
Query Match:	88.10%	Indels:	0
DB:	7	Gaps:	0

US-09-858-332G-15 (1-12) x CF965114 (1-437)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        182 CACCACCACGACCACAACCACAACCACAACCACAAC 217
```

RESULT 200
CF986640
LOCUS CF986640 438 bp mRNA linear EST 04-FEB-2005
DEFINITION 23319rsicee_2448.y1 Oryza sativa cv. LYP9 tillering whole plant cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA sequence.
ACCESSION CF986640
VERSION CF986640.1 GI:58588332
KEYWORDS EST.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 438)
AUTHORS Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W., Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X., Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J., Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q., Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z., Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P., Wu,S. and Liu,J.
TITLE The Genomes of Oryza sativa: A History of Duplications
JOURNAL PLoS Biol. 3 (2), e38 (2005)
PUBMED 15685292
COMMENT Contact: Yan Zhou
Bioinformatics Department
Hangzhou Genomics Institute
No.51 Zhijiang Road, Hangzhou 310008, China
Tel: 86-571-56805886
Fax: 86-571-56805884
Email: zhouyan@genomics.org.cn
Seq primer: M13 Forward
High quality sequence stop: 438
POLYA=No.

FEATURES Location/Qualifiers
source 1. .438
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/cultivar="LYP9"
/db_xref="taxon:39946"
/tissue_type="whole plant"
/dev_stage="tillering"
/clone_lib="Oryza sativa cv. LYP9 tillering whole plant cDNA library"

ORIGIN

Alignment Scores:

Pred. No.:	1.63	Length:	438
Score:	74.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	83.33%	Mismatches:	0
Query Match:	88.10%	Indels:	0
DB:	7	Gaps:	0

US-09-858-332G-15 (1-12) x CF986640 (1-438)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        207 CACCACCACGACCACAACCACAACCACAACCACAAC 242
```

RESULT 201
CK080515
LOCUS CK080515 438 bp mRNA linear EST 05-FEB-2005
DEFINITION 81076rsicen_7204.y1 Oryza sativa cv. LYP9 booting whole plant cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA sequence.
ACCESSION CK080515
VERSION CK080515.1 GI:58691828
KEYWORDS EST.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 438)
AUTHORS Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W., Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X., Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J., Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q., Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z., Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P., Wu,S. and Liu,J.
TITLE The Genomes of Oryza sativa: A History of Duplications
JOURNAL PLoS Biol. 3 (2), e38 (2005)
PUBMED 15685292
COMMENT Contact: Yan Zhou
Bioinformatics Department
Hangzhou Genomics Institute
No.51 Zhijiang Road, Hangzhou 310008, China
Tel: 86-571-56805886
Fax: 86-571-56805884
Email: zhouyan@genomics.org.cn
Seq primer: M13 Forward
High quality sequence stop: 438
POLYA=No.

FEATURES Location/Qualifiers
source 1. .438
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/cultivar="LYP9"
/db_xref="taxon:39946"

/tissue_type="whole plant"
/dev_stage="booting"
/clone_lib="Oryza sativa cv. LYP9 booting whole plant cDNA library"
ORIGIN

Alignment Scores:
Pred. No.: 1.63 Length: 438
Score: 74.00 Matches: 10
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 83.33% Mismatches: 0
Query Match: 88.10% Indels: 0
DB: 7 Gaps: 0

US-09-858-332G-15 (1-12) x CK080515 (1-438)

Qy 1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
||||:::|||:::||||||||||||||||||||||||
Db 192 CACCACCACGACCACAACCACAACCACAACCACAAC 227

RESULT 202
AW596891
LOCUS AW596891 440 bp mRNA linear EST 14-JUL-2004
DEFINITION sj84b04.y1 Gm-c1034 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1034-1640 5' similar to SW:IF4V_TOBAC Q40465 EUKARYOTIC INITIATION FACTOR 4A-11 ;, mRNA sequence.
ACCESSION AW596891
VERSION AW596891.1 GI:7284297
KEYWORDS EST.
SOURCE Glycine max (soybean)
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE 1 (bases 1 to 440)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
When it has been determined, an EST from the other end of this clone is listed in the 'Other ESTs on clone' field. This clone is available through: Biogenetic Services, 801 32nd Ave. Brookings, SD 57006 USA (phone: 800 423 4163; email: info@biogeneticservices.com)
Insert Length: 1578 Std Error: 0.00.
FEATURES Location/Qualifiers
source 1. .440
/organism="Glycine max"
/mol_type="mRNA"
/cultivar="Williams"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1034-1640"
/tissue_type="very young seed coats of greenhouse grown plants"
/lab_host="DH10B"
/clone_lib="Gm-c1034"
/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; This cDNA library was constructed from mRNA isolated from very young seed coats (20-50mgs fresh weight) of greenhouse grown plants. The library was prepared using the Life Technologies pSuperScript cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly(dT) sequence with a NotI restrictions site. SalI linkers adapters were ligated to the blunt-ended cDNA fragments followed by NotI digestion. The cDNA fragments were directionally cloned into the NotI-SalI restriction site of the pSPORT1 vector. The library was transformed into E.coli ElectroMax DH10B host cells. This library was constructed by Dr. Lila Vodkin and Dr. Anu Khanna."
ORIGIN

Alignment Scores:
Pred. No.: 1.64 Length: 440
Score: 74.00 Matches: 10
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 83.33% Mismatches: 0
Query Match: 88.10% Indels: 0
DB: 1 Gaps: 0

US-09-858-332G-15 (1-12) x AW596891 (1-440)

Qy 1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
||||:::|||:::||||||||||||||||||||||||
Db 28 CACAGCCACAGCCACAACCACAACCACAACCACAAC 63

RESULT 203
CF965703
LOCUS CF965703 448 bp mRNA linear EST 15-JAN-2004
DEFINITION 14418rsiced_19030.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA sequence.
ACCESSION CF965703
VERSION CF965703.1 GI:38481614
KEYWORDS EST.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 448)
AUTHORS Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W., Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X., Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J., Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q., Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z., Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P., Wu,S. and Liu,J.
TITLE The Genomes of Oryza sativa: A History of Duplications
JOURNAL PLoS Biol. 3 (2), e38 (2005)
PUBMED 15685292
COMMENT Contact: Yan Zhou
Bioinformatics Department
Hangzhou Genomics Institute
No.51 Zhijiang Road, Hangzhou 310008, China
Tel: 86-571-56805886
Fax: 86-571-56805884
Email: zhouyan@genomics.org.cn
Seq primer: M13 Forward
High quality sequence stop: 448
POLYA=No.
FEATURES Location/Qualifiers
source 1. .448
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/cultivar="LYP9"
/db_xref="taxon:39946"
/tissue_type="whole plant"
/dev_stage="trefoil"
/clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA library"
ORIGIN

Alignment Scores:
Pred. No.: 1.67 Length: 448

```
Score:                  74.00          Matches:        10
Percent Similarity:     100.00%        Conservative:   2
Best Local Similarity:  83.33%         Mismatches:     0
Query Match:            88.10%         Indels:         0
DB:                     7              Gaps:           0

US-09-858-332G-15 (1-12) x CF965703 (1-448)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        118 CACCACCACGACCACAACCACAACCACAACCACAAC 153

RESULT 204
CF960519
LOCUS       CF960519                449 bp    mRNA    linear   EST 15-JAN-2004
DEFINITION  12399rsiced_7134.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF960519
VERSION     CF960519.1  GI:38476386
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 449
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .449
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="trefoil"
                     /clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA
                     library"
ORIGIN

Alignment Scores:
Pred. No.:              1.68           Length:         449
Score:                  74.00          Matches:        10
Percent Similarity:     100.00%        Conservative:   2
Best Local Similarity:  83.33%         Mismatches:     0
Query Match:            88.10%         Indels:         0
DB:                     7              Gaps:           0

US-09-858-332G-15 (1-12) x CF960519 (1-449)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        188 CACCACCACGACCACAACCACAACCACAACCACAAC 223

RESULT 205
CK746439/c
LOCUS       CK746439                449 bp    mRNA    linear   EST 07-JUN-2005
DEFINITION  wmi01-8ms1-e03 Wmi01 Welwitschia mirabilis cDNA clone
            wmi01-8ms1-e03 5', mRNA sequence.
ACCESSION   CK746439
VERSION     CK746439.1  GI:42636862
KEYWORDS    EST.
SOURCE      Welwitschia mirabilis
  ORGANISM  Welwitschia mirabilis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Gnetophyta; Gnetopsida; Welwitschiales;
            Welwitschiaceae; Welwitschia.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   dePamphilis,C., Carlson,J., Ma,H., Soltis,D., Soltis,P.,
            Oppenheimer,D., Frohlich,M., Doyle,J., Tanksley,S., Webb,M.,
            Leebens-Mack,J., Landherr,L., Ilut,D. and Wall,K.
  TITLE     Generation of ESTs from young male strobilus of Welwitschia
            mirabilis
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Claude dePamphilis or James Leebens-Mack
            Mueller Laboratory
            Penn State University
            208 Mueller Laboratory, Department of Biology, ATTN Rm212, Penn
            State University, University Park, PA 16802, USA
            Tel: 814 863 6413
            Fax: 814 865 9131
            Email: cwd3@psu.edu or jhl10@psu.edu
            The sequence provided is trimmed of vector and low quality regions.
            Full sequence and original trace file are available from the Plant
            Genome Network website (http://pgn.cornell.edu)
            Plate: wmi01-8ms1  row: e  column: 03
            Seq primer: M13F.
FEATURES             Location/Qualifiers
     source          1. .449
                     /organism="Welwitschia mirabilis"
                     /mol_type="mRNA"
                     /db_xref="PGN:wmi01-8ms1-e03"
                     /db_xref="taxon:3377"
                     /clone="wmi01-8ms1-e03"
                     /tissue_type="young male strobilus"
                     /dev_stage="<3mm"
                     /lab_host="SOLR"
                     /clone_lib="Wmi01"
                     /note="Vector: pBluescript SK (+/-); Site_1: EcoRI;
                     Site_2: Xho1; This is a directionally cloned,
                     non-normalized library. This library has been generated by
                     the Floral Genome Project (FGP). The Floral Genome Project
                     is funded by NSF's Plant Genome Research Program
                     (DBI-0115684). More information about the project can be
                     obtained at http://fgp.bio.psu.edu"
ORIGIN

Alignment Scores:
Pred. No.:              1.68           Length:         449
Score:                  74.00          Matches:        10
Percent Similarity:     100.00%        Conservative:   2
Best Local Similarity:  83.33%         Mismatches:     0
Query Match:            88.10%         Indels:         0
DB:                     7              Gaps:           0

US-09-858-332G-15 (1-12) x CK746439 (1-449)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||||||||:::|||||||||:::|||||||||||||
Db         43 CATAATCATCATCATAATCATCATCATAATCATAAT 8

RESULT 206
CF197290
LOCUS       CF197290                460 bp    mRNA    linear   EST 31-JUL-2003
DEFINITION  Ir64vi2.1_C07 IR64-VI Oryza sativa (indica cultivar-group) cDNA
            clone IR64-VI2.1C07 5', mRNA sequence.
ACCESSION   CF197290
```

```
VERSION     CF197290.1  GI:33381947
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 460)
  AUTHORS   Ventelon-Debout,M., Thi-Thu Huang,N., Wissocq,A., Berger,C.,
            Michele,L., Piegu,B., Cooke,R., Ghesquiere,A., Delseny,M. and
            Brugidou,C.
  TITLE     Metabolism and Photosynthesis gene expression is affected in
            response to rice yellow mottle virus infection in Oryza sativa
            indica and japonica culltivars
  JOURNAL   Mol. Genet. Genomics (2003) In press
COMMENT     Contact: Brugidou C
            IRD
            BP 64501, 34394 Montpellier cedex 5 France
            Tel: 33 4 67 41 67 39
            Fax: 33 4 67 41 61 81
            PCR PRimers
            FORWARD: 5'AATTAACCCTCACTAAAGGG3'
            BACKWARD: 5'CGGGATATCACTCAGCATAATG3'
            Plate: 2.1  row: C  column: 7.
FEATURES             Location/Qualifiers
     source          1. .460
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="IR64"
                     /db_xref="taxon:39946"
                     /clone="IR64-VI2.1C07"
                     /tissue_type="leaves"
                     /dev_stage="3 weeks-old plants"
                     /lab_host="SOLR strain (Stratagene)"
                     /clone_lib="IR64-VI"
                     /note="Vector: pBluescript SK (+/-); Site_1: Eco RI;
                     Site_2: Xho I; Sample name: IR64-VI; Plant growth place:
                     greenhouse; Soil conditions: standard; Sowing date:
                     28/02/2000; Harvesting date: 17/03/2000; Stress date:
                     15/03/2000; Cultivar: ir64; Description: IR64-RYMV
                     inoculated leaves cDNA library; experimental condition:
                     rymv inoculation;"
ORIGIN

Alignment Scores:
Pred. No.:              1.72          Length:         460
Score:                  74.00         Matches:        10
Percent Similarity:     100.00%       Conservative:   2
Best Local Similarity:  83.33%        Mismatches:     0
Query Match:            88.10%        Indels:         0
DB:                     6             Gaps:           0

US-09-858-332G-15 (1-12) x CF197290 (1-460)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        190 CACCACCACGACCACAACCACAACCACAACCACAAC 225

RESULT 207
CF985613
LOCUS       CF985613                 471 bp    mRNA    linear   EST 04-FEB-2005
DEFINITION  17031rsicee_14712.y1 Oryza sativa cv. LYP9 tillering whole plant
            cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF985613
VERSION     CF985613.1  GI:58587305
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 471)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 471
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .471
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="tillering"
                     /clone_lib="Oryza sativa cv. LYP9 tillering whole plant
                     cDNA library"
ORIGIN

Alignment Scores:
Pred. No.:              1.76          Length:         471
Score:                  74.00         Matches:        10
Percent Similarity:     100.00%       Conservative:   2
Best Local Similarity:  83.33%        Mismatches:     0
Query Match:            88.10%        Indels:         0
DB:                     7             Gaps:           0

US-09-858-332G-15 (1-12) x CF985613 (1-471)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        193 CACCACCACGACCACAACCACAACCACAACCACAAC 228

RESULT 208
CF962395
LOCUS       CF962395                 475 bp    mRNA    linear   EST 15-JAN-2004
DEFINITION  9808siced_0882.z1 Oryza sativa cv. LYP9 trefoil whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF962395
VERSION     CF962395.1  GI:38478262
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 475)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
```

```
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 475
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .475
                      /organism="Oryza sativa (indica cultivar-group)"
                      /mol_type="mRNA"
                      /cultivar="LYP9"
                      /db_xref="taxon:39946"
                      /tissue_type="whole plant"
                      /dev_stage="trefoil"
                      /clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA
                      library"
ORIGIN

Alignment Scores:
Pred. No.:                1.78          Length:          475
Score:                    74.00         Matches:         10
Percent Similarity:       100.00%       Conservative:    2
Best Local Similarity:    83.33%        Mismatches:      0
Query Match:              88.10%        Indels:          0
DB:                       7             Gaps:            0

US-09-858-332G-15 (1-12) x CF962395 (1-475)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::||||||||||||||||||||||||
Db        183 CACCACCACGACCACAACCACAACCACAACCACAAC 218

RESULT 209
CF961968
LOCUS       CF961968                476 bp    mRNA    linear   EST 15-JAN-2004
DEFINITION  13885rsiced_16815.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF961968
VERSION     CF961968.1  GI:38477835
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 476)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 476
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .476
                      /organism="Oryza sativa (indica cultivar-group)"
                      /mol_type="mRNA"
                      /cultivar="LYP9"
                      /db_xref="taxon:39946"
                      /tissue_type="whole plant"
                      /dev_stage="trefoil"
                      /clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA
                      library"
ORIGIN

Alignment Scores:
Pred. No.:                1.78          Length:          476
Score:                    74.00         Matches:         10
Percent Similarity:       100.00%       Conservative:    2
Best Local Similarity:    83.33%        Mismatches:      0
Query Match:              88.10%        Indels:          0
DB:                       7             Gaps:            0

US-09-858-332G-15 (1-12) x CF961968 (1-476)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::||||||||||||||||||||||||
Db        200 CACCACCACGACCACAACCACAACCACAACCACAAC 235

RESULT 210
AQ659741/c
LOCUS       AQ659741                476 bp    DNA     linear   GSS 23-JUN-1999
DEFINITION  Sheared DNA-16J1.TR Sheared DNA Trypanosoma brucei genomic clone
            Sheared DNA-16J1, genomic survey sequence.
ACCESSION   AQ659741
VERSION     AQ659741.1  GI:5167509
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 476)
  AUTHORS   El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
            Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S.,
            Donelson,J., Fraser,C. and Adams,M.
  TITLE     Determination of clone end sequences from Trypanosoma brucei GUTat
            10.1 sheared DNA library
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: Sheared DNA-16J1.TF
            Contact: Najib M. El-Sayed
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: nelsayed@tigr.org
            Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
            DNA library constructed at TIGR. Clones will be available for
            distribution through ATCC. Sheared DNA end sequences search page:
            http://www.tigr.org/tdb/mdb/tbdb/.
            Seq primer: M13-Reverse
            Class: shotgun.
FEATURES              Location/Qualifiers
     source           1. .476
                      /organism="Trypanosoma brucei"
                      /mol_type="genomic DNA"
                      /strain="TREU927/4 GUTat 10.1"
                      /db_xref="taxon:5691"
                      /clone="Sheared DNA-16J1"
                      /clone_lib="Sheared DNA"
                      /note="Vector: pUC18; Site_1: SmaI; Constructed at The
                      Institute for Genomic Research (TIGR), Rockville, MD.
                      Genomic DNA isolated from a cloned population of
                      Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
                      sheared to give a tight size distribution (approx 2 kb).
                      The v + i method used for the library construction is
                      described in detail in Smith, H.O. and Venter, J.C.
                      (Making small insert libraries for whole genome shotgun
                      sequencing projects.  In Genome Sequencing: A Practical
                      Approach, eds. M. Vaudin and B. Barell, Oxford University
```

```
                    Press, 1999)."
ORIGIN

Alignment Scores:
Pred. No.:              1.78          Length:         476
Score:                  74.00         Matches:        10
Percent Similarity:     100.00%       Conservative:   2
Best Local Similarity:  83.33%        Mismatches:     0
Query Match:            88.10%        Indels:         0
DB:                     9             Gaps:           0

US-09-858-332G-15 (1-12) x AQ659741 (1-476)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        272 CACCACCACCACCACAATCACAACCACAACCACAAC 237

RESULT 211
CF961117
LOCUS       CF961117                486 bp    mRNA    linear   EST 15-JAN-2004
DEFINITION  11158rsiced_2453.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF961117
VERSION     CF961117.1  GI:38476984
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 486)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 486
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .486
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="trefoil"
                     /clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA
                     library"
ORIGIN

Alignment Scores:
Pred. No.:              1.82          Length:         486
Score:                  74.00         Matches:        10
Percent Similarity:     100.00%       Conservative:   2
Best Local Similarity:  83.33%        Mismatches:     0
Query Match:            88.10%        Indels:         0
DB:                     7             Gaps:           0

US-09-858-332G-15 (1-12) x CF961117 (1-486)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        166 CACCACCACGACCACAACCACAACCACAACCACAAC 201

RESULT 212
CF963249
LOCUS       CF963249                491 bp    mRNA    linear   EST 15-JAN-2004
DEFINITION  12143rsiced_6341.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF963249
VERSION     CF963249.1  GI:38479116
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 491)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 491
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .491
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="trefoil"
                     /clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA
                     library"
ORIGIN

Alignment Scores:
Pred. No.:              1.84          Length:         491
Score:                  74.00         Matches:        10
Percent Similarity:     100.00%       Conservative:   2
Best Local Similarity:  83.33%        Mismatches:     0
Query Match:            88.10%        Indels:         0
DB:                     7             Gaps:           0

US-09-858-332G-15 (1-12) x CF963249 (1-491)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        160 CACCACCACGACCACAACCACAACCACAACCACAAC 195

RESULT 213
CB964758/c
LOCUS       CB964758                492 bp    mRNA    linear   EST 29-APR-2003
DEFINITION  NL11_D05 Drought stress (leaf) Oryza sativa (indica cultivar-group)
            cDNA clone NL11_D05 3', mRNA sequence.
ACCESSION   CB964758
VERSION     CB964758.1  GI:30226866
```

```
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 492)
  AUTHORS   Markandeya,G., Ravindra Babu,P., Venkat Reddy,B., Nagabhushana,I.,
            Chandra Sekhar,A., Bennetzen,J.L., Ramakrishna,W. and Reddy,A.R.
  TITLE     ESTs from a normalized cDNA library of drought stressed rice
            seedlings (Oryza sativa L.cv Nagina 22)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Reddy AR
            Department of Plant Sciences, School of Life Sciences
            University of Hyderabad
            P.O. Central University, Hyderabad-500 046, A.P, India
            Tel: 0091-40-3010265
            Fax: 0091-40-3010145
            Email: arjulsl@uohyd.ernet.in
            Insert Length: 492   Std Error: 0.00
            Seq primer: CGCCAGGGTTTTCCCAGTCACGAC.
FEATURES            Location/Qualifiers
     source         1. .492
                    /organism="Oryza sativa (indica cultivar-group)"
                    /mol_type="mRNA"
                    /cultivar="Nagina 22"
                    /db_xref="taxon:39946"
                    /clone="NL11_D05"
                    /tissue_type="Entire leaf tissue"
                    /dev_stage="35 day-old seedlings"
                    /clone_lib="Drought stress (leaf)"
                    /note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized
                    leaf cDNA Library from drought stressed seedlings "
ORIGIN

Alignment Scores:
Pred. No.:                1.84          Length:          492
Score:                    74.00         Matches:         10
Percent Similarity:       100.00%       Conservative:    2
Best Local Similarity:    83.33%        Mismatches:      0
Query Match:              88.10%        Indels:          0
DB:                       6             Gaps:            0

US-09-858-332G-15 (1-12) x CB964758 (1-492)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::||||||||||||||||||||||||
Db        373 CACCACCACGACCACAACCACAACCACAACCACAAC 338

RESULT 214
CF960382
LOCUS       CF960382                 495 bp    mRNA    linear   EST 15-JAN-2004
DEFINITION  12469rsiced_7450.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF960382
VERSION     CF960382.1  GI:38476249
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 495)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 495
            POLYA=No.
FEATURES            Location/Qualifiers
     source         1. .495
                    /organism="Oryza sativa (indica cultivar-group)"
                    /mol_type="mRNA"
                    /cultivar="LYP9"
                    /db_xref="taxon:39946"
                    /tissue_type="whole plant"
                    /dev_stage="trefoil"
                    /clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA
                    library"
ORIGIN

Alignment Scores:
Pred. No.:                1.86          Length:          495
Score:                    74.00         Matches:         10
Percent Similarity:       100.00%       Conservative:    2
Best Local Similarity:    83.33%        Mismatches:      0
Query Match:              88.10%        Indels:          0
DB:                       7             Gaps:            0

US-09-858-332G-15 (1-12) x CF960382 (1-495)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::||||||||||||||||||||||||
Db        120 CACCACCACGACCACAACCACAACCACAACCACAAC 155

RESULT 215
CF988679
LOCUS       CF988679                 497 bp    mRNA    linear   EST 04-FEB-2005
DEFINITION  17338rsicee_15193.y1 Oryza sativa cv. LYP9 tillering whole plant
            cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF988679
VERSION     CF988679.1  GI:58590371
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 497)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 497
            POLYA=No.
FEATURES            Location/Qualifiers
```

```
     source          1. .497
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="tillering"
                     /clone_lib="Oryza sativa cv. LYP9 tillering whole plant
                     cDNA library"
ORIGIN

Alignment Scores:
Pred. No.:              1.86          Length:        497
Score:                  74.00         Matches:       10
Percent Similarity:     100.00%       Conservative:  2
Best Local Similarity:  83.33%        Mismatches:    0
Query Match:            88.10%        Indels:        0
DB:                     7             Gaps:          0

US-09-858-332G-15 (1-12) x CF988679 (1-497)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        175 CACCACCACGACCACAACCACAACCACAACCACAAC 210

RESULT 216
CF991700
LOCUS       CF991700                499 bp    mRNA    linear   EST 04-FEB-2005
DEFINITION  16342rsicee_13610.y1 Oryza sativa cv. LYP9 tillering whole plant
            cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF991700
VERSION     CF991700.1  GI:58593392
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 499)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 499
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .499
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="tillering"
                     /clone_lib="Oryza sativa cv. LYP9 tillering whole plant
                     cDNA library"
ORIGIN

Alignment Scores:
Pred. No.:              1.87          Length:        499
Score:                  74.00         Matches:       10
Percent Similarity:     100.00%       Conservative:  2
Best Local Similarity:  83.33%        Mismatches:    0
Query Match:            88.10%        Indels:        0
DB:                     7             Gaps:          0

US-09-858-332G-15 (1-12) x CF991700 (1-499)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        195 CACCACCACGACCACAACCACAACCACAACCACAAC 230

RESULT 217
CF963876
LOCUS       CF963876                505 bp    mRNA    linear   EST 15-JAN-2004
DEFINITION  14637rsiced_19454.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF963876
VERSION     CF963876.1  GI:38479787
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 505)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 505
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .505
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="trefoil"
                     /clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA
                     library"
ORIGIN

Alignment Scores:
Pred. No.:              1.89          Length:        505
Score:                  74.00         Matches:       10
Percent Similarity:     100.00%       Conservative:  2
Best Local Similarity:  83.33%        Mismatches:    0
Query Match:            88.10%        Indels:        0
DB:                     7             Gaps:          0

US-09-858-332G-15 (1-12) x CF963876 (1-505)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        201 CACCACCACGACCACAACCACAACCACAACCACAAC 236
```

RESULT 218
CF965313
LOCUS CF965313 514 bp mRNA linear EST 15-JAN-2004
DEFINITION 11490rsiced_3177.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA sequence.
ACCESSION CF965313
VERSION CF965313.1 GI:38481224
KEYWORDS EST.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 514)
AUTHORS Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W., Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X., Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J., Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q., Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z., Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P., Wu,S. and Liu,J.
TITLE The Genomes of Oryza sativa: A History of Duplications
JOURNAL PLoS Biol. 3 (2), e38 (2005)
PUBMED 15685292
COMMENT Contact: Yan Zhou
Bioinformatics Department
Hangzhou Genomics Institute
No.51 Zhijiang Road, Hangzhou 310008, China
Tel: 86-571-56805886
Fax: 86-571-56805884
Email: zhouyan@genomics.org.cn
Seq primer: M13 Forward
High quality sequence stop: 514
POLYA=No.
FEATURES Location/Qualifiers
source 1. .514
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/cultivar="LYP9"
/db_xref="taxon:39946"
/tissue_type="whole plant"
/dev_stage="trefoil"
/clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA library"
ORIGIN

Alignment Scores:

Pred. No.:	1.93	Length:	514
Score:	74.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	83.33%	Mismatches:	0
Query Match:	88.10%	Indels:	0
DB:	7	Gaps:	0

US-09-858-332G-15 (1-12) x CF965313 (1-514)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        133 CACCACCACGACCACAACCACAACCACAACCACAAC 168
```

RESULT 219
CF985124
LOCUS CF985124 516 bp mRNA linear EST 04-FEB-2005
DEFINITION 20265rsicee_1641.y1 Oryza sativa cv. LYP9 tillering whole plant cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA sequence.
ACCESSION CF985124
VERSION CF985124.1 GI:58586816
KEYWORDS EST.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 516)
AUTHORS Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W., Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X., Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J., Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q., Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z., Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P., Wu,S. and Liu,J.
TITLE The Genomes of Oryza sativa: A History of Duplications
JOURNAL PLoS Biol. 3 (2), e38 (2005)
PUBMED 15685292
COMMENT Contact: Yan Zhou
Bioinformatics Department
Hangzhou Genomics Institute
No.51 Zhijiang Road, Hangzhou 310008, China
Tel: 86-571-56805886
Fax: 86-571-56805884
Email: zhouyan@genomics.org.cn
Seq primer: M13 Forward
High quality sequence stop: 516
POLYA=No.
FEATURES Location/Qualifiers
source 1. .516
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/cultivar="LYP9"
/db_xref="taxon:39946"
/tissue_type="whole plant"
/dev_stage="tillering"
/clone_lib="Oryza sativa cv. LYP9 tillering whole plant cDNA library"
ORIGIN

Alignment Scores:

Pred. No.:	1.94	Length:	516
Score:	74.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	83.33%	Mismatches:	0
Query Match:	88.10%	Indels:	0
DB:	7	Gaps:	0

US-09-858-332G-15 (1-12) x CF985124 (1-516)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        173 CACCACCACGACCACAACCACAACCACAACCACAAC 208
```

RESULT 220
BU673005/c
LOCUS BU673005 518 bp mRNA linear EST 07-OCT-2002
DEFINITION NL_9_53 Drought stress (leaf) Oryza sativa (indica cultivar-group) cDNA clone NL_9_53 3', mRNA sequence.
ACCESSION BU673005
VERSION BU673005.1 GI:23513147
KEYWORDS EST.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 518)
AUTHORS Reddy,A.R., Markandeya,G., Ramakrishna,W., Nagabhushana,I., Ravindra Babu,P., Madana Mohan Reddy,A., Chandra Sekhar,A. and Bennetzen,J.L.
TITLE Novel EST enrichment with normalized cDNA libraries from drought stressed rice (Oryza sativa L.cv Nagina 22) (2002)
JOURNAL Unpublished (2002)
COMMENT Contact: Reddy AR
Department of Plant Sciences, School of Life Sciences
University of Hyderabad

```
           P.O. Central University, Hyderabad-500 046, A.P, India
           Tel: 0091-40-3010265
           Fax: 0091-40-3010145
           Email: arjulsl@uohyd.ernet.in
           Insert Length: 518   Std Error: 0.00
           Seq primer: CGCCAGGGTTTTCCCAGTCACGAC.
FEATURES             Location/Qualifiers
     source          1. .518
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nagina 22"
                     /db_xref="taxon:39946"
                     /clone="NL_9_53"
                     /tissue_type="Entire leaf tissue"
                     /dev_stage="35 day-old seedlings"
                     /clone_lib="Drought stress (leaf)"
                     /note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized
                     leaf cDNA Library from drought stressed seedlings "
ORIGIN

Alignment Scores:
Pred. No.:              1.95          Length:         518
Score:                  74.00         Matches:        10
Percent Similarity:     100.00%       Conservative:   2
Best Local Similarity:  83.33%        Mismatches:     0
Query Match:            88.10%        Indels:         0
DB:                     5             Gaps:           0

US-09-858-332G-15 (1-12) x BU673005 (1-518)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        372 CACCACCACGACCACAACCACAACCACAACCACAAC 337

RESULT 221
BU673252/c
LOCUS       BU673252                518 bp    mRNA    linear   EST 07-OCT-2002
DEFINITION  NL_13_3 Drought stress (leaf) Oryza sativa (indica cultivar-group)
            cDNA clone NL_13_3 3', mRNA sequence.
ACCESSION   BU673252
VERSION     BU673252.1  GI:23513679
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 518)
  AUTHORS   Reddy,A.R., Markandeya,G., Ramakrishna,W., Nagabhushana,I.,
            Ravindra Babu,P., Madana Mohan Reddy,A., Chandra Sekhar,A. and
            Bennetzen,J.L.
  TITLE     Novel EST enrichment with normalized cDNA libraries from drought
            stressed rice (Oryza sativa L.cv Nagina 22) (2002)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Reddy AR
            Department of Plant Sciences, School of Life Sciences
            University of Hyderabad
            P.O. Central University, Hyderabad-500 046, A.P, India
            Tel: 0091-40-3010265
            Fax: 0091-40-3010145
            Email: arjulsl@uohyd.ernet.in
            Insert Length: 518   Std Error: 0.00
            Seq primer: CGCCAGGGTTTTCCCAGTCACGAC.
FEATURES             Location/Qualifiers
     source          1. .518
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nagina 22"
                     /db_xref="taxon:39946"
                     /clone="NL_13_3"
                     /tissue_type="Entire leaf tissue"
                     /dev_stage="35 day-old seedlings"
                     /clone_lib="Drought stress (leaf)"
                     /note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized
                     leaf cDNA Library from drought stressed seedlings "
ORIGIN

Alignment Scores:
Pred. No.:              1.95          Length:         518
Score:                  74.00         Matches:        10
Percent Similarity:     100.00%       Conservative:   2
Best Local Similarity:  83.33%        Mismatches:     0
Query Match:            88.10%        Indels:         0
DB:                     5             Gaps:           0

US-09-858-332G-15 (1-12) x BU673252 (1-518)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        372 CACCACCACGACCACAACCACAACCACAACCACAAC 337

RESULT 222
CB965151/c
LOCUS       CB965151                518 bp    mRNA    linear   EST 29-APR-2003
DEFINITION  NL17_E02 Drought stress (leaf) Oryza sativa (indica cultivar-group)
            cDNA clone NL17_E02 3', mRNA sequence.
ACCESSION   CB965151
VERSION     CB965151.1  GI:30227259
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 518)
  AUTHORS   Markandeya,G., Ravindra Babu,P., Venkat Reddy,B., Nagabhushana,I.,
            Chandra Sekhar,A., Bennetzen,J.L., Ramakrishna,W. and Reddy,A.R.
  TITLE     ESTs from a normalized cDNA library of drought stressed rice
            seedlings (Oryza sativa L.cv Nagina 22)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Reddy AR
            Department of Plant Sciences, School of Life Sciences
            University of Hyderabad
            P.O. Central University, Hyderabad-500 046, A.P, India
            Tel: 0091-40-3010265
            Fax: 0091-40-3010145
            Email: arjulsl@uohyd.ernet.in
            Insert Length: 518   Std Error: 0.00
            Seq primer: CGCCAGGGTTTTCCCAGTCACGAC.
FEATURES             Location/Qualifiers
     source          1. .518
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nagina 22"
                     /db_xref="taxon:39946"
                     /clone="NL17_E02"
                     /tissue_type="Entire leaf tissue"
                     /dev_stage="35 day-old seedlings"
                     /clone_lib="Drought stress (leaf)"
                     /note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized
                     leaf cDNA Library from drought stressed seedlings "
ORIGIN

Alignment Scores:
Pred. No.:              1.95          Length:         518
Score:                  74.00         Matches:        10
Percent Similarity:     100.00%       Conservative:   2
Best Local Similarity:  83.33%        Mismatches:     0
Query Match:            88.10%        Indels:         0
DB:                     6             Gaps:           0

US-09-858-332G-15 (1-12) x CB965151 (1-518)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        372 CACCACCACGACCACAACCACAACCACAACCACAAC 337
```

```
RESULT 223
CF964678
LOCUS       CF964678                 521 bp    mRNA    linear   EST 15-JAN-2004
DEFINITION  12289rsiced_6750.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF964678
VERSION     CF964678.1  GI:38480589
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 521)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
  PUBMED    15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 521
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .521
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="trefoil"
                     /clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA
                     library"
ORIGIN

Alignment Scores:
Pred. No.:                1.96          Length:          521
Score:                    74.00         Matches:         10
Percent Similarity:       100.00%       Conservative:    2
Best Local Similarity:    83.33%        Mismatches:      0
Query Match:              88.10%        Indels:          0
DB:                       7             Gaps:            0

US-09-858-332G-15 (1-12) x CF964678 (1-521)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        200 CACCACCACGACCACAACCACAACCACAACCACAAC 235

RESULT 224
CF985622
LOCUS       CF985622                 529 bp    mRNA    linear   EST 04-FEB-2005
DEFINITION  17719rsicee_8347.y1 Oryza sativa cv. LYP9 tillering whole plant
            cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF985622
VERSION     CF985622.1  GI:58587314
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 529)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
  PUBMED    15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 529
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .529
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="tillering"
                     /clone_lib="Oryza sativa cv. LYP9 tillering whole plant
                     cDNA library"
ORIGIN

Alignment Scores:
Pred. No.:                1.99          Length:          529
Score:                    74.00         Matches:         10
Percent Similarity:       100.00%       Conservative:    2
Best Local Similarity:    83.33%        Mismatches:      0
Query Match:              88.10%        Indels:          0
DB:                       7             Gaps:            0

US-09-858-332G-15 (1-12) x CF985622 (1-529)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        195 CACCACCACGACCACAACCACAACCACAACCACAAC 230

RESULT 225
CB965275/c
LOCUS       CB965275                 531 bp    mRNA    linear   EST 29-APR-2003
DEFINITION  NL19_A06 Drought stress (leaf) Oryza sativa (indica cultivar-group)
            cDNA clone NL19_A06 3', mRNA sequence.
ACCESSION   CB965275
VERSION     CB965275.1  GI:30227383
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 531)
  AUTHORS   Markandeya,G., Ravindra Babu,P., Venkat Reddy,B., Nagabhushana,I.,
            Chandra Sekhar,A., Bennetzen,J.L., Ramakrishna,W. and Reddy,A.R.
  TITLE     ESTs from a normalized cDNA library of drought stressed rice
            seedlings (Oryza sativa L.cv Nagina 22)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Reddy AR
            Department of Plant Sciences, School of Life Sciences
            University of Hyderabad
            P.O. Central University, Hyderabad-500 046, A.P, India
```

```
           Tel: 0091-40-3010265
           Fax: 0091-40-3010145
           Email: arjulsl@uohyd.ernet.in
           Insert Length: 531   Std Error: 0.00
           Seq primer: CGCCAGGGTTTTCCCAGTCACGAC.
FEATURES             Location/Qualifiers
     source          1. .531
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nagina 22"
                     /db_xref="taxon:39946"
                     /clone="NL19_A06"
                     /tissue_type="Entire leaf tissue"
                     /dev_stage="35 day-old seedlings"
                     /clone_lib="Drought stress (leaf)"
                     /note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized
                     leaf cDNA Library from drought stressed seedlings "
ORIGIN

Alignment Scores:
Pred. No.:              2             Length:        531
Score:                  74.00         Matches:       10
Percent Similarity:     100.00%       Conservative:  2
Best Local Similarity:  83.33%        Mismatches:    0
Query Match:            88.10%        Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-15 (1-12) x CB965275 (1-531)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        372 CACCACCACGACCACAACCACAACCACAACCACAAC 337

RESULT 226
CB967418/c
LOCUS       CB967418                531 bp    mRNA    linear   EST 29-APR-2003
DEFINITION  NL52_B12 Drought stress (leaf) Oryza sativa (indica cultivar-group)
            cDNA clone NL52_B12 3', mRNA sequence.
ACCESSION   CB967418
VERSION     CB967418.1  GI:30229527
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 531)
  AUTHORS   Markandeya,G., Ravindra Babu,P., Venkat Reddy,B., Nagabhushana,I.,
            Chandra Sekhar,A., Bennetzen,J.L., Ramakrishna,W. and Reddy,A.R.
  TITLE     ESTs from a normalized cDNA library of drought stressed rice
            seedlings (Oryza sativa L.cv Nagina 22)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Reddy AR
            Department of Plant Sciences, School of Life Sciences
            University of Hyderabad
            P.O. Central University, Hyderabad-500 046, A.P, India
            Tel: 0091-40-3010265
            Fax: 0091-40-3010145
            Email: arjulsl@uohyd.ernet.in
            Insert Length: 531   Std Error: 0.00
            Seq primer: CGCCAGGGTTTTCCCAGTCACGAC.
FEATURES             Location/Qualifiers
     source          1. .531
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nagina 22"
                     /db_xref="taxon:39946"
                     /clone="NL52_B12"
                     /tissue_type="Entire leaf tissue"
                     /dev_stage="35 day-old seedlings"
                     /clone_lib="Drought stress (leaf)"
                     /note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized
                     leaf cDNA Library from drought stressed seedlings "
ORIGIN

Alignment Scores:
Pred. No.:              2             Length:        531
Score:                  74.00         Matches:       10
Percent Similarity:     100.00%       Conservative:  2
Best Local Similarity:  83.33%        Mismatches:    0
Query Match:            88.10%        Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-15 (1-12) x CB967418 (1-531)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        369 CACCACCACGACCACAACCACAACCACAACCACAAC 334

RESULT 227
CB967427/c
LOCUS       CB967427                531 bp    mRNA    linear   EST 29-APR-2003
DEFINITION  NL52_D11 Drought stress (leaf) Oryza sativa (indica cultivar-group)
            cDNA clone NL52_D11 3', mRNA sequence.
ACCESSION   CB967427
VERSION     CB967427.1  GI:30229536
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 531)
  AUTHORS   Markandeya,G., Ravindra Babu,P., Venkat Reddy,B., Nagabhushana,I.,
            Chandra Sekhar,A., Bennetzen,J.L., Ramakrishna,W. and Reddy,A.R.
  TITLE     ESTs from a normalized cDNA library of drought stressed rice
            seedlings (Oryza sativa L.cv Nagina 22)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Reddy AR
            Department of Plant Sciences, School of Life Sciences
            University of Hyderabad
            P.O. Central University, Hyderabad-500 046, A.P, India
            Tel: 0091-40-3010265
            Fax: 0091-40-3010145
            Email: arjulsl@uohyd.ernet.in
            Insert Length: 531   Std Error: 0.00
            Seq primer: CGCCAGGGTTTTCCCAGTCACGAC.
FEATURES             Location/Qualifiers
     source          1. .531
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nagina 22"
                     /db_xref="taxon:39946"
                     /clone="NL52_D11"
                     /tissue_type="Entire leaf tissue"
                     /dev_stage="35 day-old seedlings"
                     /clone_lib="Drought stress (leaf)"
                     /note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized
                     leaf cDNA Library from drought stressed seedlings "
ORIGIN

Alignment Scores:
Pred. No.:              2             Length:        531
Score:                  74.00         Matches:       10
Percent Similarity:     100.00%       Conservative:  2
Best Local Similarity:  83.33%        Mismatches:    0
Query Match:            88.10%        Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-15 (1-12) x CB967427 (1-531)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        369 CACCACCACGACCACAACCACAACCACAACCACAAC 334

RESULT 228
```

```
BI798956
LOCUS       BI798956                532 bp    mRNA    linear   EST 02-OCT-2001
DEFINITION  H121D04 Endosperm library from Oryza sativa (10 days after
            anthesis) Oryza sativa cDNA clone H121D04, mRNA sequence.
ACCESSION   BI798956
VERSION     BI798956.1  GI:15850680
KEYWORDS    EST.
SOURCE      Oryza sativa
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 532)
  AUTHORS   Dong,H.T., Li,D.B., Zhuang,X.F., Dai,C.G., Sun,L.X., Pei,Y.X.,
            Wu,H.F., Jiang,Y.X., Yu,F.C., Gao,Q.K. and Lou,Y.C.
  TITLE     A Gene Expression Screen in Oryza sativa
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Haitao Dong, Debao Li
            Bioinfomatics and Gene Network Research Group
            Zhejiang University
            Kaixuan Road 268#, Hangzhou, Zhejiang, P.R.China
            Tel: 0086-571-86892051
            Fax: 0086-571-86961525
            Email: webmaster@estarray.org, URL: http://www.estarray.org
            Seq primer: M13 forward primer.
FEATURES             Location/Qualifiers
     source          1. .532
                     /organism="Oryza sativa"
                     /mol_type="mRNA"
                     /db_xref="taxon:4530"
                     /clone="H121D04"
                     /tissue_type="Endosperm"
                     /dev_stage="10 days after anthesis"
                     /clone_lib="Endosperm library from Oryza sativa (10 days
                     after anthesis)"
                     /note="Vector: pSport2"
ORIGIN

Alignment Scores:
Pred. No.:                2              Length:         532
Score:                    74.00          Matches:        10
Percent Similarity:       100.00%        Conservative:   2
Best Local Similarity:    83.33%         Mismatches:     0
Query Match:              88.10%         Indels:         0
DB:                       3              Gaps:           0

US-09-858-332G-15 (1-12) x BI798956 (1-532)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        143 CACCACCACGACCACAACCACAACCACAACCACAAC 178

RESULT 229
CK082135
LOCUS       CK082135                537 bp    mRNA    linear   EST 05-FEB-2005
DEFINITION  80962rsicen_6962.y1 Oryza sativa cv. LYP9 booting whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CK082135
VERSION     CK082135.1  GI:58693448
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 537)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
  PUBMED    15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 537
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .537
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="booting"
                     /clone_lib="Oryza sativa cv. LYP9 booting whole plant cDNA
                     library"
ORIGIN

Alignment Scores:
Pred. No.:                2.02           Length:         537
Score:                    74.00          Matches:        10
Percent Similarity:       100.00%        Conservative:   2
Best Local Similarity:    83.33%         Mismatches:     0
Query Match:              88.10%         Indels:         0
DB:                       7              Gaps:           0

US-09-858-332G-15 (1-12) x CK082135 (1-537)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        148 CACCACCACGACCACAACCACAACCACAACCACAAC 183

RESULT 230
CF960953
LOCUS       CF960953                542 bp    mRNA    linear   EST 15-JAN-2004
DEFINITION  8110rsiced_11548.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF960953
VERSION     CF960953.1  GI:38476820
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 542)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
  PUBMED    15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
```

```
          High quality sequence stop: 542
          POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .542
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="trefoil"
                     /clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA
                     library"
ORIGIN

Alignment Scores:
Pred. No.:              2.04          Length:        542
Score:                  74.00         Matches:       10
Percent Similarity:     100.00%       Conservative:  2
Best Local Similarity:  83.33%        Mismatches:    0
Query Match:            88.10%        Indels:        0
DB:                     7             Gaps:          0

US-09-858-332G-15 (1-12) x CF960953 (1-542)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        190 CACCACCACGACCACAACCACAACCACAACCACAAC 225

RESULT 231
CF966553
LOCUS       CF966553                542 bp    mRNA    linear   EST 15-JAN-2004
DEFINITION  10790rsiced_14486.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF966553
VERSION     CF966553.1  GI:38482524
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 542)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 542
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .542
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="trefoil"
                     /clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA
                     library"
ORIGIN

Alignment Scores:
Pred. No.:              2.04          Length:        542
Score:                  74.00         Matches:       10
Percent Similarity:     100.00%       Conservative:  2
Best Local Similarity:  83.33%        Mismatches:    0
Query Match:            88.10%        Indels:        0
DB:                     7             Gaps:          0

US-09-858-332G-15 (1-12) x CF966553 (1-542)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        164 CACCACCACGACCACAACCACAACCACAACCACAAC 199

RESULT 232
CF966531
LOCUS       CF966531                543 bp    mRNA    linear   EST 15-JAN-2004
DEFINITION  7915rsiced_11317.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF966531
VERSION     CF966531.1  GI:38482497
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 543)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 543
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .543
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="trefoil"
                     /clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA
                     library"
ORIGIN

Alignment Scores:
Pred. No.:              2.04          Length:        543
Score:                  74.00         Matches:       10
Percent Similarity:     100.00%       Conservative:  2
Best Local Similarity:  83.33%        Mismatches:    0
Query Match:            88.10%        Indels:        0
DB:                     7             Gaps:          0

US-09-858-332G-15 (1-12) x CF966531 (1-543)
```

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||||||
Db        179 CACCACCACGACCACAACCACAACCACAACCACAAC 214

RESULT 233
CF991407
LOCUS       CF991407                 543 bp    mRNA    linear   EST 04-FEB-2005
DEFINITION  18863rsicee_10294.y1 Oryza sativa cv. LYP9 tillering whole plant
            cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF991407
VERSION     CF991407.1  GI:58593099
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 543)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
  PUBMED    15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 543
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .543
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="tillering"
                     /clone_lib="Oryza sativa cv. LYP9 tillering whole plant
                     cDNA library"
ORIGIN

Alignment Scores:
Pred. No.:                2.04          Length:          543
Score:                    74.00         Matches:         10
Percent Similarity:       100.00%       Conservative:    2
Best Local Similarity:    83.33%        Mismatches:      0
Query Match:              88.10%        Indels:          0
DB:                       7             Gaps:            0

US-09-858-332G-15 (1-12) x CF991407 (1-543)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||||||
Db        152 CACCACCACGACCACAACCACAACCACAACCACAAC 187

RESULT 234
CF965538
LOCUS       CF965538                 546 bp    mRNA    linear   EST 15-JAN-2004
DEFINITION  7941rsiced_11345.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF965538
VERSION     CF965538.1  GI:38481449
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 546)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
  PUBMED    15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 546
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .546
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="trefoil"
                     /clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA
                     library"
ORIGIN

Alignment Scores:
Pred. No.:                2.06          Length:          546
Score:                    74.00         Matches:         10
Percent Similarity:       100.00%       Conservative:    2
Best Local Similarity:    83.33%        Mismatches:      0
Query Match:              88.10%        Indels:          0
DB:                       7             Gaps:            0

US-09-858-332G-15 (1-12) x CF965538 (1-546)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||||||
Db        170 CACCACCACGACCACAACCACAACCACAACCACAAC 205

RESULT 235
CF961410
LOCUS       CF961410                 548 bp    mRNA    linear   EST 15-JAN-2004
DEFINITION  12411rsiced_7148.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF961410
VERSION     CF961410.1  GI:38477277
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 548)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
```

```
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 548
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .548
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="trefoil"
                     /clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA
                     library"
ORIGIN

Alignment Scores:
Pred. No.:              2.06          Length:        548
Score:                  74.00         Matches:       10
Percent Similarity:     100.00%       Conservative:  2
Best Local Similarity:  83.33%        Mismatches:    0
Query Match:            88.10%        Indels:        0
DB:                     7             Gaps:          0

US-09-858-332G-15 (1-12) x CF961410 (1-548)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        159 CACCACCACGACCACAACCACAACCACAACCACAAC 194

RESULT 236
BX900515/c
LOCUS       BX900515                549 bp    mRNA    linear   EST 16-JAN-2004
DEFINITION  BX900515 Oryza sativa library (Han B) Oryza sativa cDNA clone
            y708b01p3, mRNA sequence.
ACCESSION   BX900515
VERSION     BX900515.1  GI:40492928
KEYWORDS    EST.
SOURCE      Oryza sativa
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 549)
  AUTHORS   Han,B., Feng,Q., Huang,Y.C., Ying,K., Li,Y., Guan,J.P., Zhu,J.J.,
            Zhao,Q., Hu,X., Liu,Y.L., Mu,J., Yu,Z., Chen,L., Fan,D.L.,
            Weng,Q.J., Zhang,L., Lu,Y.Q., Yu,S.L., Liu,X.H., Lu,T.T.,
            Zhang,Y.J., Lu,Y., Li,C., Li,T., Zhang,Y., Hu,H., Jia,P.X.,
            Zhang,L., Lan,L.F., Chen,W., Wu,S.A. and Xue,Y.B.
  TITLE     Rice cDNA EST clone
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Han Bin
            National Center for Gene Research
            Chinese Academy of Sciences
            500# Cao Bao Road, Shanghai 200233, China
            Email: bhan@ncgr.ac.cn
            Clone requests: bhan@ncgr.ac.cn
             This is rice cdna est clone
            Web site: http://www.ncgr.ac.cn.
FEATURES             Location/Qualifiers
     source          1. .549
                     /organism="Oryza sativa"
                     /mol_type="mRNA"
                     /db_xref="taxon:4530"
                     /clone="y708b01p3"
                     /clone_lib="Oryza sativa library (Han B)"
ORIGIN

Alignment Scores:
Pred. No.:              2.07          Length:        549
Score:                  74.00         Matches:       10
Percent Similarity:     100.00%       Conservative:  2
Best Local Similarity:  83.33%        Mismatches:    0
Query Match:            88.10%        Indels:        0
DB:                     5             Gaps:          0

US-09-858-332G-15 (1-12) x BX900515 (1-549)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        438 CACCACCACGACCACAACCACAACCACAACCACAAC 403

RESULT 237
CB967111/c
LOCUS       CB967111                549 bp    mRNA    linear   EST 29-APR-2003
DEFINITION  NL48_B05 Drought stress (leaf) Oryza sativa (indica cultivar-group)
            cDNA clone NL48_B05 3', mRNA sequence.
ACCESSION   CB967111
VERSION     CB967111.1  GI:30229220
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 549)
  AUTHORS   Markandeya,G., Ravindra Babu,P., Venkat Reddy,B., Nagabhushana,I.,
            Chandra Sekhar,A., Bennetzen,J.L., Ramakrishna,W. and Reddy,A.R.
  TITLE     ESTs from a normalized cDNA library of drought stressed rice
            seedlings (Oryza sativa L.cv Nagina 22)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Reddy AR
            Department of Plant Sciences, School of Life Sciences
            University of Hyderabad
            P.O. Central University, Hyderabad-500 046, A.P, India
            Tel: 0091-40-3010265
            Fax: 0091-40-3010145
            Email: arjulsl@uohyd.ernet.in
            Insert Length: 549   Std Error: 0.00
            Seq primer: CGCCAGGGTTTTCCCAGTCACGAC.
FEATURES             Location/Qualifiers
     source          1. .549
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nagina 22"
                     /db_xref="taxon:39946"
                     /clone="NL48_B05"
                     /tissue_type="Entire leaf tissue"
                     /dev_stage="35 day-old seedlings"
                     /clone_lib="Drought stress (leaf)"
                     /note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized
                     leaf cDNA Library from drought stressed seedlings "
ORIGIN

Alignment Scores:
Pred. No.:              2.07          Length:        549
Score:                  74.00         Matches:       10
Percent Similarity:     100.00%       Conservative:  2
Best Local Similarity:  83.33%        Mismatches:    0
Query Match:            88.10%        Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-15 (1-12) x CB967111 (1-549)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
```

```
                    |||:::|||:::|||||||||||||||||||||||||
Db        372 CACCACCACGACCACAACCACAACCACAACCACAAC 337

RESULT 238
CF961941
LOCUS       CF961941                549 bp    mRNA    linear   EST 15-JAN-2004
DEFINITION  13585rsiced_5272.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF961941
VERSION     CF961941.1  GI:38477808
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 549)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 549
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .549
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="trefoil"
                     /clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA
                     library"
ORIGIN

Alignment Scores:
Pred. No.:                2.07          Length:          549
Score:                    74.00         Matches:         10
Percent Similarity:       100.00%       Conservative:    2
Best Local Similarity:    83.33%        Mismatches:      0
Query Match:              88.10%        Indels:          0
DB:                       7             Gaps:            0

US-09-858-332G-15 (1-12) x CF961941 (1-549)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        177 CACCACCACGACCACAACCACAACCACAACCACAAC 212

RESULT 239
BU673593/c
LOCUS       BU673593                550 bp    mRNA    linear   EST 07-OCT-2002
DEFINITION  NL_17_56 Drought stress (leaf) Oryza sativa (indica cultivar-group)
            cDNA clone NL_17_56 3', mRNA sequence.
ACCESSION   BU673593
VERSION     BU673593.1  GI:23514344
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 550)
  AUTHORS   Reddy,A.R., Markandeya,G., Ramakrishna,W., Nagabhushana,I.,
            Ravindra Babu,P., Madana Mohan Reddy,A., Chandra Sekhar,A. and
            Bennetzen,J.L.
  TITLE     Novel EST enrichment with normalized cDNA libraries from drought
            stressed rice (Oryza sativa L.cv Nagina 22) (2002)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Reddy AR
            Department of Plant Sciences, School of Life Sciences
            University of Hyderabad
            P.O. Central University, Hyderabad-500 046, A.P, India
            Tel: 0091-40-3010265
            Fax: 0091-40-3010145
            Email: arjulsl@uohyd.ernet.in
            Insert Length: 550   Std Error: 0.00
            Seq primer: CGCCAGGGTTTTCCCAGTCACGAC.
FEATURES             Location/Qualifiers
     source          1. .550
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nagina 22"
                     /db_xref="taxon:39946"
                     /clone="NL_17_56"
                     /tissue_type="Entire leaf tissue"
                     /dev_stage="35 day-old seedlings"
                     /clone_lib="Drought stress (leaf)"
                     /note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized
                     leaf cDNA Library from drought stressed seedlings "
ORIGIN

Alignment Scores:
Pred. No.:                2.07          Length:          550
Score:                    74.00         Matches:         10
Percent Similarity:       100.00%       Conservative:    2
Best Local Similarity:    83.33%        Mismatches:      0
Query Match:              88.10%        Indels:          0
DB:                       5             Gaps:            0

US-09-858-332G-15 (1-12) x BU673593 (1-550)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        373 CACCACCACGACCACAACCACAACCACAACCACAAC 338

RESULT 240
BU673632/c
LOCUS       BU673632                550 bp    mRNA    linear   EST 07-OCT-2002
DEFINITION  NL_18_10 Drought stress (leaf) Oryza sativa (indica cultivar-group)
            cDNA clone NL_18_10 3', mRNA sequence.
ACCESSION   BU673632
VERSION     BU673632.1  GI:23514422
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 550)
  AUTHORS   Reddy,A.R., Markandeya,G., Ramakrishna,W., Nagabhushana,I.,
            Ravindra Babu,P., Madana Mohan Reddy,A., Chandra Sekhar,A. and
            Bennetzen,J.L.
  TITLE     Novel EST enrichment with normalized cDNA libraries from drought
            stressed rice (Oryza sativa L.cv Nagina 22) (2002)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Reddy AR
            Department of Plant Sciences, School of Life Sciences
            University of Hyderabad
            P.O. Central University, Hyderabad-500 046, A.P, India
            Tel: 0091-40-3010265
```

```
            Fax: 0091-40-3010145
            Email: arjulsl@uohyd.ernet.in
            Insert Length: 550   Std Error: 0.00
            Seq primer: CGCCAGGGTTTTCCCAGTCACGAC.
FEATURES              Location/Qualifiers
     source           1. .550
                      /organism="Oryza sativa (indica cultivar-group)"
                      /mol_type="mRNA"
                      /cultivar="Nagina 22"
                      /db_xref="taxon:39946"
                      /clone="NL_18_10"
                      /tissue_type="Entire leaf tissue"
                      /dev_stage="35 day-old seedlings"
                      /clone_lib="Drought stress (leaf)"
                      /note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized
                      leaf cDNA Library from drought stressed seedlings "
ORIGIN

Alignment Scores:
Pred. No.:              2.07          Length:        550
Score:                  74.00         Matches:       10
Percent Similarity:     100.00%       Conservative:  2
Best Local Similarity:  83.33%        Mismatches:    0
Query Match:            88.10%        Indels:        0
DB:                     5             Gaps:          0

US-09-858-332G-15 (1-12) x BU673632 (1-550)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        373 CACCACCACGACCACAACCACAACCACAACCACAAC 338

RESULT 241
CB964701/c
LOCUS       CB964701                550 bp    mRNA    linear   EST 29-APR-2003
DEFINITION  NL_10C08 Drought stress (leaf) Oryza sativa (indica cultivar-group)
            cDNA clone NL_10C08 3', mRNA sequence.
ACCESSION   CB964701
VERSION     CB964701.1  GI:30226809
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 550)
  AUTHORS   Markandeya,G., Ravindra Babu,P., Venkat Reddy,B., Nagabhushana,I.,
            Chandra Sekhar,A., Bennetzen,J.L., Ramakrishna,W. and Reddy,A.R.
  TITLE     ESTs from a normalized cDNA library of drought stressed rice
            seedlings (Oryza sativa L.cv Nagina 22)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Reddy AR
            Department of Plant Sciences, School of Life Sciences
            University of Hyderabad
            P.O. Central University, Hyderabad-500 046, A.P, India
            Tel: 0091-40-3010265
            Fax: 0091-40-3010145
            Email: arjulsl@uohyd.ernet.in
            Insert Length: 550   Std Error: 0.00
            Seq primer: CGCCAGGGTTTTCCCAGTCACGAC.
FEATURES              Location/Qualifiers
     source           1. .550
                      /organism="Oryza sativa (indica cultivar-group)"
                      /mol_type="mRNA"
                      /cultivar="Nagina 22"
                      /db_xref="taxon:39946"
                      /clone="NL_10C08"
                      /tissue_type="Entire leaf tissue"
                      /dev_stage="35 day-old seedlings"
                      /clone_lib="Drought stress (leaf)"
                      /note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized
                      leaf cDNA Library from drought stressed seedlings "
ORIGIN

Alignment Scores:
Pred. No.:              2.07          Length:        550
Score:                  74.00         Matches:       10
Percent Similarity:     100.00%       Conservative:  2
Best Local Similarity:  83.33%        Mismatches:    0
Query Match:            88.10%        Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-15 (1-12) x CB964701 (1-550)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        373 CACCACCACGACCACAACCACAACCACAACCACAAC 338

RESULT 242
CB966849/c
LOCUS       CB966849                550 bp    mRNA    linear   EST 29-APR-2003
DEFINITION  NL43_F12 Drought stress (leaf) Oryza sativa (indica cultivar-group)
            cDNA clone NL43_F12 3', mRNA sequence.
ACCESSION   CB966849
VERSION     CB966849.1  GI:30228958
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 550)
  AUTHORS   Markandeya,G., Ravindra Babu,P., Venkat Reddy,B., Nagabhushana,I.,
            Chandra Sekhar,A., Bennetzen,J.L., Ramakrishna,W. and Reddy,A.R.
  TITLE     ESTs from a normalized cDNA library of drought stressed rice
            seedlings (Oryza sativa L.cv Nagina 22)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Reddy AR
            Department of Plant Sciences, School of Life Sciences
            University of Hyderabad
            P.O. Central University, Hyderabad-500 046, A.P, India
            Tel: 0091-40-3010265
            Fax: 0091-40-3010145
            Email: arjulsl@uohyd.ernet.in
            Insert Length: 550   Std Error: 0.00
            Seq primer: CGCCAGGGTTTTCCCAGTCACGAC.
FEATURES              Location/Qualifiers
     source           1. .550
                      /organism="Oryza sativa (indica cultivar-group)"
                      /mol_type="mRNA"
                      /cultivar="Nagina 22"
                      /db_xref="taxon:39946"
                      /clone="NL43_F12"
                      /tissue_type="Entire leaf tissue"
                      /dev_stage="35 day-old seedlings"
                      /clone_lib="Drought stress (leaf)"
                      /note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized
                      leaf cDNA Library from drought stressed seedlings "
ORIGIN

Alignment Scores:
Pred. No.:              2.07          Length:        550
Score:                  74.00         Matches:       10
Percent Similarity:     100.00%       Conservative:  2
Best Local Similarity:  83.33%        Mismatches:    0
Query Match:            88.10%        Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-15 (1-12) x CB966849 (1-550)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        373 CACCACCACGACCACAACCACAACCACAACCACAAC 338

RESULT 243
CK758412/c
```

```
LOCUS       CK758412                555 bp    mRNA    linear   EST 09-JUN-2005
DEFINITION  wmi01-10ms2-e04 Wmi01 Welwitschia mirabilis cDNA clone
            wmi01-10ms2-e04 5', mRNA sequence.
ACCESSION   CK758412
VERSION     CK758412.1  GI:42648835
KEYWORDS    EST.
SOURCE      Welwitschia mirabilis
  ORGANISM  Welwitschia mirabilis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Gnetophyta; Gnetopsida; Welwitschiales;
            Welwitschiaceae; Welwitschia.
REFERENCE   1  (bases 1 to 555)
  AUTHORS   dePamphilis,C., Carlson,J., Ma,H., Soltis,D., Soltis,P.,
            Oppenheimer,D., Frohlich,M., Doyle,J., Tanksley,S., Webb,M.,
            Leebens-Mack,J., Landherr,L., Ilut,D. and Wall,K.
  TITLE     Generation of ESTs from young male strobilus of Welwitschia
            mirabilis
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Claude dePamphilis or James Leebens-Mack
            Mueller Laboratory
            Penn State University
            208 Mueller Laboratory, Department of Biology, ATTN Rm212, Penn
            State University, University Park, PA 16802, USA
            Tel: 814 863 6413
            Fax: 814 865 9131
            Email: cwd3@psu.edu or jhl10@psu.edu
            The sequence provided is trimmed of vector and low quality regions.
            Full sequence and original trace file are available from the Plant
            Genome Network website (http://pgn.cornell.edu)
            Plate: wmi01-10ms2  row: e  column: 04
            Seq primer: M13F.
FEATURES             Location/Qualifiers
     source          1. .555
                     /organism="Welwitschia mirabilis"
                     /mol_type="mRNA"
                     /db_xref="PGN:wmi01-10ms2-e04"
                     /db_xref="taxon:3377"
                     /clone="wmi01-10ms2-e04"
                     /tissue_type="young male strobilus"
                     /dev_stage="<3mm"
                     /lab_host="SOLR"
                     /clone_lib="Wmi01"
                     /note="Vector: pBluescript SK (+/-); Site_1: EcoRI;
                     Site_2: XhoI; This is a directionally cloned,
                     non-normalized library. This library has been generated by
                     the Floral Genome Project (FGP). The Floral Genome Project
                     is funded by NSF's Plant Genome Research Program
                     (DBI-0115684). More information about the project can be
                     obtained at http://fgp.bio.psu.edu"
ORIGIN

Alignment Scores:
Pred. No.:                2.09          Length:          555
Score:                    74.00         Matches:         10
Percent Similarity:       100.00%       Conservative:    2
Best Local Similarity:    83.33%        Mismatches:      0
Query Match:              88.10%        Indels:          0
DB:                       7             Gaps:            0

US-09-858-332G-15 (1-12) x CK758412 (1-555)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||||||||:::|||||||||:::||||||||||||
Db        221 CATAATCATCATCATAATCATCATCATAATCATAAT 186

RESULT 244
CK766845/c
LOCUS       CK766845                560 bp    mRNA    linear   EST 08-JUN-2005
DEFINITION  wmi01-6ms3-g08 Wmi01 Welwitschia mirabilis cDNA clone
            wmi01-6ms3-g08 5', mRNA sequence.
ACCESSION   CK766845
VERSION     CK766845.1  GI:42720744
KEYWORDS    EST.
SOURCE      Welwitschia mirabilis
  ORGANISM  Welwitschia mirabilis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Gnetophyta; Gnetopsida; Welwitschiales;
            Welwitschiaceae; Welwitschia.
REFERENCE   1  (bases 1 to 560)
  AUTHORS   dePamphilis,C., Carlson,J., Ma,H., Soltis,D., Soltis,P.,
            Oppenheimer,D., Frohlich,M., Doyle,J., Tanksley,S., Webb,M.,
            Leebens-Mack,J., Landherr,L., Ilut,D. and Wall,K.
  TITLE     Generation of ESTs from young male strobilus of Welwitschia
            mirabilis
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Claude dePamphilis or James Leebens-Mack
            Mueller Laboratory
            Penn State University
            208 Mueller Laboratory, Department of Biology, ATTN Rm212, Penn
            State University, University Park, PA 16802, USA
            Tel: 814 863 6413
            Fax: 814 865 9131
            Email: cwd3@psu.edu or jhl10@psu.edu
            The sequence provided is trimmed of vector and low quality regions.
            Full sequence and original trace file are available from the Plant
            Genome Network website (http://pgn.cornell.edu)
            Plate: wmi01-6ms3  row: g  column: 08
            Seq primer: M13F.
FEATURES             Location/Qualifiers
     source          1. .560
                     /organism="Welwitschia mirabilis"
                     /mol_type="mRNA"
                     /db_xref="PGN:wmi01-6ms3-g08"
                     /db_xref="taxon:3377"
                     /clone="wmi01-6ms3-g08"
                     /tissue_type="young male strobilus"
                     /dev_stage="<3mm"
                     /lab_host="SOLR"
                     /clone_lib="Wmi01"
                     /note="Vector: pBluescript SK (+/-); Site_1: EcoRI;
                     Site_2: XhoI; This is a directionally cloned,
                     non-normalized library. This library has been generated by
                     the Floral Genome Project (FGP). The Floral Genome Project
                     is funded by NSF's Plant Genome Research Program
                     (DBI-0115684). More information about the project can be
                     obtained at http://fgp.bio.psu.edu"
ORIGIN

Alignment Scores:
Pred. No.:                2.11          Length:          560
Score:                    74.00         Matches:         10
Percent Similarity:       100.00%       Conservative:    2
Best Local Similarity:    83.33%        Mismatches:      0
Query Match:              88.10%        Indels:          0
DB:                       7             Gaps:            0

US-09-858-332G-15 (1-12) x CK766845 (1-560)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||||||||:::|||||||||:::||||||||||||
Db        293 CATAATCATCATCATAATCATCATCATAATCATAAT 258

RESULT 245
CF963651
LOCUS       CF963651                562 bp    mRNA    linear   EST 15-JAN-2004
DEFINITION  8463rsiced_12091.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF963651
VERSION     CF963651.1  GI:38479548
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
```

```
REFERENCE   1  (bases 1 to 562)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 562
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .562
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="trefoil"
                     /clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA
                     library"
ORIGIN

Alignment Scores:
Pred. No.:                2.12          Length:          562
Score:                    74.00         Matches:         10
Percent Similarity:       100.00%       Conservative:    2
Best Local Similarity:    83.33%        Mismatches:      0
Query Match:              88.10%        Indels:          0
DB:                       7             Gaps:            0

US-09-858-332G-15 (1-12) x CF963651 (1-562)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        187 CACCACCACGACCACAACCACAACCACAACCACAAC 222

RESULT 246
CF966423
LOCUS       CF966423                563 bp    mRNA    linear   EST 15-JAN-2004
DEFINITION  9528rsiced_9721.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF966423
VERSION     CF966423.1  GI:38482382
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 563)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 563
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .563
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="trefoil"
                     /clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA
                     library"
ORIGIN

Alignment Scores:
Pred. No.:                2.12          Length:          563
Score:                    74.00         Matches:         10
Percent Similarity:       100.00%       Conservative:    2
Best Local Similarity:    83.33%        Mismatches:      0
Query Match:              88.10%        Indels:          0
DB:                       7             Gaps:            0

US-09-858-332G-15 (1-12) x CF966423 (1-563)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        198 CACCACCACGACCACAACCACAACCACAACCACAAC 233

RESULT 247
CF961387
LOCUS       CF961387                564 bp    mRNA    linear   EST 15-JAN-2004
DEFINITION  13695rsiced_4146.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF961387
VERSION     CF961387.1  GI:38477254
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 564)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 564
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .564
                     /organism="Oryza sativa (indica cultivar-group)"
```

/mol_type="mRNA"
/cultivar="LYP9"
/db_xref="taxon:39946"
/tissue_type="whole plant"
/dev_stage="trefoil"
/clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA library"
ORIGIN

Alignment Scores:

Pred. No.:	2.13	Length:	564
Score:	74.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	83.33%	Mismatches:	0
Query Match:	88.10%	Indels:	0
DB:	7	Gaps:	0

US-09-858-332G-15 (1-12) x CF961387 (1-564)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::||||||||||||||||||||||||
Db        206 CACCACCACGACCACAACCACAACCACAACCACAAC 241
```

RESULT 248
BM143854
LOCUS BM143854 565 bp mRNA linear EST 16-JUL-2004
DEFINITION saj49e11.y1 Gm-c1072 Glycine max cDNA clone SOYBEAN CLONE ID: Gm-c1072-2926 5' similar to TR:O23660 O23660 HYPOTHETICAL PROTEIN. ;, mRNA sequence.
ACCESSION BM143854
VERSION BM143854.1 GI:17153921
KEYWORDS EST.
SOURCE Glycine max (soybean)
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE 1 (bases 1 to 565)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
When it has been determined, an EST from the other end of this clone is listed in the 'Other ESTs on clone' field. This clone is available through: Biogenetic Services, 801 32nd Ave. Brookings, SD 57006 USA (phone: 800 423 4163; email: info@biogeneticservices.com)
Seq primer: -40RP from Gibco
High quality sequence stop: 422.
FEATURES Location/Qualifiers
source 1. .565
/organism="Glycine max"
/mol_type="mRNA"
/cultivar="PI567374"
/db_xref="taxon:3847"
/clone="SOYBEAN CLONE ID: Gm-c1072-2926"
/tissue_type="seedlings induced for symptoms of SDS (Sudden Death Syndrome) disease"
/dev_stage="2-3 weeks old"
/lab_host="DH10B"
/clone_lib="Gm-c1072"
/note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2: XhoI; The cDNA library wms constructed from mRNA isolated from 2-3 week old seedlings that were induced for symptoms of SDS (Sudden Death Syndrome) disease by the translocation of culture filtrate of Fusarium solani f. sp. glycines (Plant Cell Report 18:375-380). Cultivar PI 567374 is partially resistant to the disease SDS. Plant tissue (expanded leaves, folded leaves, and new shoots) were collected at 1, 6, 24, and 48 hrs. after inoculation and their mRNA pooled equally for cDNA construction. The library was prepared using the Stratagene pBluescript II SK(+) library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with an XhoI restriction site. EcoRI adaptors were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA insert is protected from XhoI digestion via methylation during first strand synthesis. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into E.coli ElectroMax DH10B host cells. Plants were inoculated by Shuxian Li (Glen Hartman lab, University of Illinois). Library was constructed by Steve Clough (Lila Vodkin lab, University of Illinois)."
ORIGIN

Alignment Scores:

Pred. No.:	2.13	Length:	565
Score:	74.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	83.33%	Mismatches:	0
Query Match:	88.10%	Indels:	0
DB:	3	Gaps:	0

US-09-858-332G-15 (1-12) x BM143854 (1-565)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||||||||||||||:::||||||||||||
Db        114 CACAGCCACAATCACAACCACCACCACAACCATAAC 149
```

RESULT 249
CF959693
LOCUS CF959693 565 bp mRNA linear EST 15-JAN-2004
DEFINITION 7904rsiced_11304.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA sequence.
ACCESSION CF959693
VERSION CF959693.1 GI:38475560
KEYWORDS EST.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 565)
AUTHORS Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W., Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X., Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J., Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q., Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z., Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P., Wu,S. and Liu,J.
TITLE The Genomes of Oryza sativa: A History of Duplications
JOURNAL PLoS Biol. 3 (2), e38 (2005)
PUBMED 15685292
COMMENT Contact: Yan Zhou
Bioinformatics Department
Hangzhou Genomics Institute
No.51 Zhijiang Road, Hangzhou 310008, China
Tel: 86-571-56805886
Fax: 86-571-56805884
Email: zhouyan@genomics.org.cn
Seq primer: M13 Forward
High quality sequence stop: 565

POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .565
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/cultivar="LYP9"
/db_xref="taxon:39946"
/tissue_type="whole plant"
/dev_stage="trefoil"
/clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA library"
ORIGIN

Alignment Scores:

Pred. No.:	2.13	Length:	565
Score:	74.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	83.33%	Mismatches:	0
Query Match:	88.10%	Indels:	0
DB:	7	Gaps:	0

US-09-858-332G-15 (1-12) x CF959693 (1-565)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::||||||||||||||||||||||||
Db        176 CACCACCACGACCACAACCACAACCACAACCACAAC 211
```

RESULT 250
CF965105
LOCUS CF965105 565 bp mRNA linear EST 15-JAN-2004
DEFINITION 13519rsiced_5175.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA sequence.
ACCESSION CF965105
VERSION CF965105.1 GI:38481016
KEYWORDS EST.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 565)
AUTHORS Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W., Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X., Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J., Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q., Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z., Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P., Wu,S. and Liu,J.
TITLE The Genomes of Oryza sativa: A History of Duplications
JOURNAL PLoS Biol. 3 (2), e38 (2005)
PUBMED 15685292
COMMENT Contact: Yan Zhou
Bioinformatics Department
Hangzhou Genomics Institute
No.51 Zhijiang Road, Hangzhou 310008, China
Tel: 86-571-56805886
Fax: 86-571-56805884
Email: zhouyan@genomics.org.cn
Seq primer: M13 Forward
High quality sequence stop: 565
POLYA=No.
FEATURES Location/Qualifiers
source 1. .565
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/cultivar="LYP9"
/db_xref="taxon:39946"
/tissue_type="whole plant"
/dev_stage="trefoil"
/clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA library"
ORIGIN

Alignment Scores:

Pred. No.:	2.13	Length:	565
Score:	74.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	83.33%	Mismatches:	0
Query Match:	88.10%	Indels:	0
DB:	7	Gaps:	0

US-09-858-332G-15 (1-12) x CF965105 (1-565)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::||||||||||||||||||||||||
Db        200 CACCACCACGACCACAACCACAACCACAACCACAAC 235
```

RESULT 251
CA486806
LOCUS CA486806 567 bp mRNA linear EST 14-NOV-2002
DEFINITION WHE4336_B08_D16ZS Wheat meiotic anther cDNA library Triticum aestivum cDNA clone WHE4336_B08_D16, mRNA sequence.
ACCESSION CA486806
VERSION CA486806.1 GI:24980878
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 567)
AUTHORS Anderson,O.D., Chao,S., Crossman,C., Langridge,P., Lazo,G.R., Pham,J., Rausch,C.J., Sutton,T., Woo,J. and Wilson,C.
TITLE The structure and function of the expressed portion of the wheat genomes - Meiotic anther cDNA library
JOURNAL Unpublished (2002)
COMMENT Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oandersn@pw.usda.gov
Sequences have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
Seq primer: SK primer.
FEATURES Location/Qualifiers
source 1. .567
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE4336_B08_D16"
/tissue_type="Anther"
/dev_stage="Meiotic stages pre-meiosis-metaphase I"
/lab_host="E. coli DH10B"
/clone_lib="Wheat meiotic anther cDNA library"
/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; Plants were grown in a glasshouse. Anther meiotic stage was determined by removing anthers from individual primary florets. One anther was sacrificed for microscopic staging, and if determined to be between (and including) meiotic stages pre-meiosis and metaphase I, the remaining two anthers were collected and pooled for library construction. The tissue, total RNA, and poly(A) RNA were prepared, cDNA synthesised, and directionally ligated into pSPORT1 by Tim Sutton in the P Langridge Lab at the Department of Plant Science, University of Adelaide, Waite Campus, Australia. Average insert size 1.5Kb. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."
ORIGIN

Alignment Scores:

Pred. No.:	2.14	Length:	567

```
Score:                 74.00          Matches:        10
Percent Similarity:    100.00%        Conservative:   2
Best Local Similarity: 83.33%         Mismatches:     0
Query Match:           88.10%         Indels:         0
DB:                    6              Gaps:           0

US-09-858-332G-15 (1-12) x CA486806 (1-567)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::||||||||||||||||||||||||
Db        180 CACCACCACGACCACAACCACAACCACAACCACAAC 215

RESULT 252
CD491431/c
LOCUS       CD491431                 570 bp    mRNA    linear   EST 05-JUN-2003
DEFINITION  WHE3087_E06_I11ZT CS wheat cold-stressed seedling subtracted cDNA
            library Triticum aestivum cDNA clone WHE3087_E06_I11, mRNA
            sequence.
ACCESSION   CD491431
VERSION     CD491431.1  GI:31419031
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 570)
  AUTHORS   Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Crossman,C.,
            Fenton,R.D., Lazo,G.R., Nguyen,H.T., Pham,J., Rausch,C.J.,
            Wilson,C., Woo,J. and Zhang,D.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - CS cold-stressed seedling subtracted cDNA library
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            This EST was generated by sequencing from the 3' end of the clone.
            Sequences have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20.
            Seq primer: T7 primer.
FEATURES             Location/Qualifiers
     source          1. .570
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /cultivar="Chinese Spring"
                     /db_xref="taxon:4565"
                     /clone="WHE3087_E06_I11"
                     /tissue_type="Seedling"
                     /dev_stage="Five-day old seedling"
                     /lab_host="E. coli SOLR"
                     /clone_lib="CS wheat cold-stressed seedling subtracted
                     cDNA library"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; Seeds were
                     surface-sterilized, germinated and grown aseptically in
                     the dark at room temperature on filter paper with water,
                     nystatin and cefotaxime in covered crystallization dishes.
                     Five-day old seedlings were transferred to 5 C cold room
                     and kept for 48 hr. The tissue, total RNA, and poly(A) RNA
                     were prepared, a cDNA library was made, and the cDNA
                     clones were in vivo excised to give pBluescript phagemids
                     in the TJ Close lab (Choi, Close, Fenton) at the
                     University of California, Riverside. The cDNA clones were
                     in vivo excised to give pBluescript phagemids before
                     subtraction was carried out. The mass excision of phagemid
                     library and subtraction were done in HT Nguyen lab by D.
                     Zhang at Texas Tech Univeristy. Plasmid DNA preparations
                     and DNA sequencing were performed in the OD Anderson lab
                     (all other authors)."
ORIGIN

Alignment Scores:
Pred. No.:             2.15           Length:         570
Score:                 74.00          Matches:        10
Percent Similarity:    100.00%        Conservative:   2
Best Local Similarity: 83.33%         Mismatches:     0
Query Match:           88.10%         Indels:         0
DB:                    6              Gaps:           0

US-09-858-332G-15 (1-12) x CD491431 (1-570)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::||||||||||||||||||||||||
Db        376 CACCACCACGACCACAACCACAACCACAACCACAAC 341

RESULT 253
BE195039
LOCUS       BE195039                 578 bp    mRNA    linear   EST 22-OCT-2001
DEFINITION  HVSMEh0088C04f Hordeum vulgare 5-45 DAP spike EST library
            HVcDNA0009 (5 to 45 DAP) Hordeum vulgare subsp. vulgare cDNA clone
            HVSMEh0088C04f, mRNA sequence.
ACCESSION   BE195039
VERSION     BE195039.2  GI:13187888
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 578)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D.,
            Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W.,
            Fenton,R.D., Close,S.J., Oates,R. and Main,D.
  TITLE     Development of a genetically and physically anchored EST resource
            for barley genomics: Morex 5-45 DAP spike cDNA library
  JOURNAL   Unpublished (2001)
COMMENT     On Jun 26, 2000 this sequence version replaced gi:8707230.
            Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total hq bases = 357
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence stop: 422.
FEATURES             Location/Qualifiers
     source          1. .578
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Morex"
                     /sub_species="vulgare"
                     /db_xref="taxon:112509"
                     /clone="HVSMEh0088C04f"
                     /tissue_type="5-45 DAP Spike"
                     /lab_host="SOLR"
                     /clone_lib="Hordeum vulgare 5-45 DAP spike EST library
                     HVcDNA0009 (5 to 45 DAP)"
                     /note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1;
                     Plants were grown in the greenhouse at the University of
                     California, Riverside (Fenton, SJ Close, TJ Close). Whole
                     spikes with awns trimmed were collected at 5, 10, 15, 20,
                     30 and 45 DAP (Fenton). Total RNA was prepared from each
                     pool, equal quantities of all six RNA pools were combined,
                     poly(A) RNA was purified from the mixture, one primary
                     unamplified cDNA library was made, and 1 million pfu were
                     in vivo excised to give pBluescript SK(-) cDNA phagemids
                     (Choi)in the TJ Close lab at the University of California,
                     Riverside. Phagemids were plated and picked at the Clemson
                     University Genomics Institute (CUGI) (Begum, Palmer,
                     Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
```

sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

ORIGIN

Alignment Scores:

Pred. No.:	2.18	Length:	578
Score:	74.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	83.33%	Mismatches:	0
Query Match:	88.10%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-15 (1-12) x BE195039 (1-578)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||||||||:::|||:::||||||||||||||||||
Db         99 CACAACCACGACCACCACCACAACCACAACCACAAC 134
```

RESULT 254
CF959561
LOCUS CF959561 579 bp mRNA linear EST 15-JAN-2004
DEFINITION 9814siced_0888.z1 Oryza sativa cv. LYP9 trefoil whole plant cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA sequence.
ACCESSION CF959561
VERSION CF959561.1 GI:38475428
KEYWORDS EST.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 579)
AUTHORS Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W., Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X., Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J., Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q., Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z., Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P., Wu,S. and Liu,J.
TITLE The Genomes of Oryza sativa: A History of Duplications
JOURNAL PLoS Biol. 3 (2), e38 (2005)
PUBMED 15685292
COMMENT Contact: Yan Zhou
Bioinformatics Department
Hangzhou Genomics Institute
No.51 Zhijiang Road, Hangzhou 310008, China
Tel: 86-571-56805886
Fax: 86-571-56805884
Email: zhouyan@genomics.org.cn
Seq primer: M13 Forward
High quality sequence stop: 579
POLYA=Yes.

FEATURES Location/Qualifiers
source 1. .579
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/cultivar="LYP9"
/db_xref="taxon:39946"
/tissue_type="whole plant"
/dev_stage="trefoil"
/clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA library"

ORIGIN

Alignment Scores:

Pred. No.:	2.19	Length:	579
Score:	74.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	83.33%	Mismatches:	0
Query Match:	88.10%	Indels:	0
DB:	7	Gaps:	0

US-09-858-332G-15 (1-12) x CF959561 (1-579)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::||||||||||||||||||||||||
Db        185 CACCACCACGACCACAACCACAACCACAACCACAAC 220
```

RESULT 255
CF984909
LOCUS CF984909 583 bp mRNA linear EST 04-FEB-2005
DEFINITION 18592sicee_6519.z1 Oryza sativa cv. LYP9 tillering whole plant cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA sequence.
ACCESSION CF984909
VERSION CF984909.1 GI:58586601
KEYWORDS EST.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 583)
AUTHORS Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W., Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X., Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J., Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q., Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z., Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P., Wu,S. and Liu,J.
TITLE The Genomes of Oryza sativa: A History of Duplications
JOURNAL PLoS Biol. 3 (2), e38 (2005)
PUBMED 15685292
COMMENT Contact: Yan Zhou
Bioinformatics Department
Hangzhou Genomics Institute
No.51 Zhijiang Road, Hangzhou 310008, China
Tel: 86-571-56805886
Fax: 86-571-56805884
Email: zhouyan@genomics.org.cn
Seq primer: M13 Forward
High quality sequence stop: 583
POLYA=Yes.

FEATURES Location/Qualifiers
source 1. .583
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/cultivar="LYP9"
/db_xref="taxon:39946"
/tissue_type="whole plant"
/dev_stage="tillering"
/clone_lib="Oryza sativa cv. LYP9 tillering whole plant cDNA library"

ORIGIN

Alignment Scores:

Pred. No.:	2.2	Length:	583
Score:	74.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	83.33%	Mismatches:	0
Query Match:	88.10%	Indels:	0
DB:	7	Gaps:	0

US-09-858-332G-15 (1-12) x CF984909 (1-583)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        194 CACCACCACGACCACAACCACAACCACAACCACAAC 229

RESULT 256
CF991926
LOCUS       CF991926                583 bp    mRNA    linear   EST 04-FEB-2005
DEFINITION  16250rsicee_12302.y1 Oryza sativa cv. LYP9 tillering whole plant
            cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF991926
VERSION     CF991926.1  GI:58593618
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 583)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
  PUBMED    15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 583
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .583
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="tillering"
                     /clone_lib="Oryza sativa cv. LYP9 tillering whole plant
                     cDNA library"
ORIGIN

Alignment Scores:
Pred. No.:                2.2          Length:          583
Score:                    74.00        Matches:         10
Percent Similarity:       100.00%      Conservative:    2
Best Local Similarity:    83.33%       Mismatches:      0
Query Match:              88.10%       Indels:          0
DB:                       7            Gaps:            0

US-09-858-332G-15 (1-12) x CF991926 (1-583)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        205 CACCACCACGACCACAACCACAACCACAACCACAAC 240

RESULT 257
CK082860
LOCUS       CK082860                584 bp    mRNA    linear   EST 05-FEB-2005
DEFINITION  85761rsicen_3177.y1 Oryza sativa cv. LYP9 booting whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CK082860
VERSION     CK082860.1  GI:58694173
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 584)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
  PUBMED    15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 584
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .584
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="booting"
                     /clone_lib="Oryza sativa cv. LYP9 booting whole plant cDNA
                     library"
ORIGIN

Alignment Scores:
Pred. No.:                2.21         Length:          584
Score:                    74.00        Matches:         10
Percent Similarity:       100.00%      Conservative:    2
Best Local Similarity:    83.33%       Mismatches:      0
Query Match:              88.10%       Indels:          0
DB:                       7            Gaps:            0

US-09-858-332G-15 (1-12) x CK082860 (1-584)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        195 CACCACCACGACCACAACCACAACCACAACCACAAC 230

RESULT 258
CF967065
LOCUS       CF967065                588 bp    mRNA    linear   EST 15-JAN-2004
DEFINITION  9827siced_0908.z1 Oryza sativa cv. LYP9 trefoil whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF967065
VERSION     CF967065.1  GI:38483263
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 588)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
```

```
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 588
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .588
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="trefoil"
                     /clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA
                     library"
ORIGIN

Alignment Scores:
Pred. No.:               2.22          Length:         588
Score:                   74.00         Matches:        10
Percent Similarity:      100.00%       Conservative:   2
Best Local Similarity:   83.33%        Mismatches:     0
Query Match:             88.10%        Indels:         0
DB:                      7             Gaps:           0

US-09-858-332G-15 (1-12) x CF967065 (1-588)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        217 CACCACCACGACCACAACCACAACCACAACCACAAC 252

RESULT 259
CF953321
LOCUS       CF953321                593 bp    mRNA    linear   EST 15-JAN-2004
DEFINITION  6565rsiceb_14824.y1 Oryza sativa cv. PA64s leaf cDNA library Oryza
            sativa (indica cultivar-group) cDNA 5', mRNA sequence.
ACCESSION   CF953321
VERSION     CF953321.1  GI:38469190
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 593)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 593
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .593
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="PA64s"
                     /db_xref="taxon:39946"
                     /tissue_type="leaf"
                     /dev_stage="trefoil"
                     /clone_lib="Oryza sativa cv. PA64s leaf cDNA library"
ORIGIN

Alignment Scores:
Pred. No.:               2.24          Length:         593
Score:                   74.00         Matches:        10
Percent Similarity:      100.00%       Conservative:   2
Best Local Similarity:   83.33%        Mismatches:     0
Query Match:             88.10%        Indels:         0
DB:                      7             Gaps:           0

US-09-858-332G-15 (1-12) x CF953321 (1-593)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        220 CACCACCACGACCACAACCACAACCACAACCACAAC 255

RESULT 260
CK079978
LOCUS       CK079978                594 bp    mRNA    linear   EST 05-FEB-2005
DEFINITION  79610rsicen_3547.y1 Oryza sativa cv. LYP9 booting whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CK079978
VERSION     CK079978.1  GI:58691291
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 594)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 594
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .594
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="booting"
                     /clone_lib="Oryza sativa cv. LYP9 booting whole plant cDNA
                     library"
ORIGIN
```

```
Alignment Scores:
Pred. No.:                2.24          Length:          594
Score:                    74.00         Matches:         10
Percent Similarity:       100.00%       Conservative:    2
Best Local Similarity:    83.33%        Mismatches:      0
Query Match:              88.10%        Indels:          0
DB:                       7             Gaps:            0

US-09-858-332G-15 (1-12) x CK079978 (1-594)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::||||||||||||||||||||||||
Db        213 CACCACCACGACCACAACCACAACCACAACCACAAC 248

RESULT 261
CZ563660
LOCUS       CZ563660                601 bp    DNA     linear   GSS 17-JUN-2005
DEFINITION  1_p_bW038J11Q_J15_10 WAG Gallus gallus genomic clone WAG-38J11,
            genomic survey sequence.
ACCESSION   CZ563660
VERSION     CZ563660.1  GI:67957372
KEYWORDS    GSS.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 601)
  AUTHORS   Crooijmans,R.P.M.A. and Groenen,M.A.M.
  TITLE     Gallus gallus WAG WL library BAC end reads
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Crooijmans RPMA & Groenen MAM
            Animal Breeding and Genetics Group
            Wageningen University
            Maijkeweg 40, 6709 PG, Wageningen, The Netherlands
            Tel: 31 317 483192; 31 317481317
            Fax: 31 317 483929
            Email: richard.crooijmans@wur.nl; martien.groenen@wur.nl
            Seq primer: universal M13Forward
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .601
                     /organism="Gallus gallus"
                     /mol_type="genomic DNA"
                     /strain="White Leghorn"
                     /db_xref="taxon:9031"
                     /clone="WAG-38J11"
                     /sex="Female"
                     /clone_lib="WAG"
                     /note="Vector: pECBAC1; cloning site HINDIII"
ORIGIN

Alignment Scores:
Pred. No.:                2.27          Length:          601
Score:                    74.00         Matches:         11
Percent Similarity:       91.67%        Conservative:    0
Best Local Similarity:    91.67%        Mismatches:      1
Query Match:              88.10%        Indels:          0
DB:                       10            Gaps:            0

US-09-858-332G-15 (1-12) x CZ563660 (1-601)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||   ||||||||||||||||||||||||||||
Db        508 CATAACCCGAACCACAACCACAACCATAACCATAAC 543

RESULT 262
BJ397270
LOCUS       BJ397270                604 bp    mRNA    linear   EST 08-MAR-2002
DEFINITION  BJ397270 Dictyostelium discoideum cDNA library, SF Dictyostelium
            discoideum cDNA clone dds45o16 5', mRNA sequence.
ACCESSION   BJ397270
VERSION     BJ397270.1  GI:19308356
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 604)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the slug stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .604
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dds45o16"
                     /sex="mat A"
                     /dev_stage="Slug stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, SF"
ORIGIN

Alignment Scores:
Pred. No.:                2.28          Length:          604
Score:                    74.00         Matches:         10
Percent Similarity:       100.00%       Conservative:    2
Best Local Similarity:    83.33%        Mismatches:      0
Query Match:              88.10%        Indels:          0
DB:                       3             Gaps:            0

US-09-858-332G-15 (1-12) x BJ397270 (1-604)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::||||||||||||||||||||||||
Db         98 CATAGTCATAGTCATAACCATAATCATAATCACAAT 133

RESULT 263
CK080511
LOCUS       CK080511                610 bp    mRNA    linear   EST 05-FEB-2005
DEFINITION  84855rsicen_26146.y1 Oryza sativa cv. LYP9 booting whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CK080511
VERSION     CK080511.1  GI:58691824
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 610)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
  PUBMED    15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
```

```
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 610
            POLYA=Yes.
FEATURES                Location/Qualifiers
     source             1. .610
                        /organism="Oryza sativa (indica cultivar-group)"
                        /mol_type="mRNA"
                        /cultivar="LYP9"
                        /db_xref="taxon:39946"
                        /tissue_type="whole plant"
                        /dev_stage="booting"
                        /clone_lib="Oryza sativa cv. LYP9 booting whole plant cDNA
                        library"
ORIGIN

Alignment Scores:
Pred. No.:              2.31            Length:         610
Score:                  74.00           Matches:        10
Percent Similarity:     100.00%         Conservative:   2
Best Local Similarity:  83.33%          Mismatches:     0
Query Match:            88.10%          Indels:         0
DB:                     7               Gaps:           0

US-09-858-332G-15 (1-12) x CK080511 (1-610)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::||||||||||||||||||||||||
Db        204 CACCACCACGACCACAACCACAACCACAACCACAAC 239

RESULT 264
CV579577
LOCUS       CV579577                611 bp    mRNA    linear   EST 22-OCT-2004
DEFINITION  kc89h11.y2 Xiphinema index CSEQDL01 Xiphinema index cDNA 5' similar
            to contains element MSR1 repetitive element ;, mRNA sequence.
ACCESSION   CV579577
VERSION     CV579577.1  GI:54550233
KEYWORDS    EST.
SOURCE      Xiphinema index
  ORGANISM  Xiphinema index
            Eukaryota; Metazoa; Nematoda; Enoplea; Dorylaimida; Dorylaimina;
            Longidoroidea; Longidoridae; Xiphinema.
REFERENCE   1  (bases 1 to 611)
  AUTHORS   McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J.,
            Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B.,
            Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C.,
            Tsagareishvili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C.,
            Underwood,K., Steptoe,M., Allen,M., Person,B., Swaller,T.,
            Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     The Washington Univ. Nematode EST Project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Mitreva M
            The Washington Univ. Nematode EST Project, 2004
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library materials provided by: J.T. Jones Library constructed by:
            Incyte Genomics (for J.T. Jones)
            Seq primer: -21UPpot primer
            High quality sequence stop: 611.
FEATURES                Location/Qualifiers
     source             1. .611
                        /organism="Xiphinema index"
                        /mol_type="mRNA"
                        /db_xref="taxon:46003"
                        /sex="mixed"
                        /dev_stage="mixed"
                        /lab_host="GC10 (only ligation mix supplied)"
                        /clone_lib="Xiphinema index CSEQDL01"
                        /note="Vector: pBluescript SK+; Site_1: 5': NotI; Site_2:
                        3': EcoRI; The library was made from mRNA isolated from
                        total RNA using oligo dT cellulose. Total RNA was
                        generated from mixed stage X. index extracted from soil
                        around fig (Ficus carica) roots. Nematodes were cleaned by
                        passing through a series of sieves.  Pure X. index were
                        obtained by hand sorting of nematodes. The library was
                        constructed by Incyte Genomics by cloning cDNA
                        directionally into in the pBluescript (SK+) vector using
                        NotI and EcoRI sites. The library was provided by John T.
                        Jones (jjones@scri.sari.ac.uk) at SCRI, Nematology
                        Department. Sequencing by: Washington University Genome
                        Sequencing Center, St. Louis, MO."
ORIGIN

Alignment Scores:
Pred. No.:              2.31            Length:         611
Score:                  74.00           Matches:        10
Percent Similarity:     100.00%         Conservative:   2
Best Local Similarity:  83.33%          Mismatches:     0
Query Match:            88.10%          Indels:         0
DB:                     7               Gaps:           0

US-09-858-332G-15 (1-12) x CV579577 (1-611)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::||||||||||||||||||||||||
Db        428 CACAGCCACAATCACAATCACAGCCACAACCACAAC 463

RESULT 265
CF963590
LOCUS       CF963590                617 bp    mRNA    linear   EST 15-JAN-2004
DEFINITION  9842siced_0927.z1 Oryza sativa cv. LYP9 trefoil whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF963590
VERSION     CF963590.1  GI:38479477
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 617)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 617
            POLYA=Yes.
FEATURES                Location/Qualifiers
     source             1. .617
                        /organism="Oryza sativa (indica cultivar-group)"
                        /mol_type="mRNA"
                        /cultivar="LYP9"
                        /db_xref="taxon:39946"
                        /tissue_type="whole plant"
                        /dev_stage="trefoil"
                        /clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA
                        library"
```

```
ORIGIN

Alignment Scores:
Pred. No.:                2.34          Length:          617
Score:                    74.00         Matches:         10
Percent Similarity:       100.00%       Conservative:    2
Best Local Similarity:    83.33%        Mismatches:      0
Query Match:              88.10%        Indels:          0
DB:                       7             Gaps:            0

US-09-858-332G-15 (1-12) x CF963590 (1-617)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::||||||||||||||||||||||||
Db        182 CACCACCACGACCACAACCACAACCACAACCACAAC 217

RESULT 266
BF108353
LOCUS       BF108353                643 bp    mRNA    linear   EST 02-FEB-2004
DEFINITION  EI31K10 Normalized rice cDNA library Oryza sativa (indica
            cultivar-group) cDNA clone EI31K10 similar to Pathogen induced
            defense-responsive sequence, mRNA sequence.
ACCESSION   BF108353
VERSION     BF108353.1  GI:40385943
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 643)
  AUTHORS   Zhou,B., Peng,K., Chu,Z., Wang,S. and Zhang,Q.
  TITLE     The defense-responsive genes showing enhanced and repressed
            expression after pathogen infection in rice (Oryza sativa L.)
  JOURNAL   Sci. China, C, Life Sci. 45 (5), 449-467 (2002)
COMMENT     Contact: Wang S
            National Key Laboratory of Crop Genetic Improvement
            Hauzhong Agricultural University
            Wuhan 430070, China
            Tel: 86-27-87282044
            Fax: 86-27-87287092
            Email: shipingwang@hotmail.com.
FEATURES             Location/Qualifiers
     source          1. .643
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Minghui 63"
                     /db_xref="taxon:39946"
                     /clone="EI31K10"
                     /tissue_type="14 different tissues"
                     /lab_host="DH10B"
                     /clone_lib="Normalized rice cDNA library"
                     /note="Vector: pSPORT 1; Site_1: Not I; Site_2: Sal I; 14
                     different tissues including callus, seedlings (etiolated,
                     three-leaf stage, and five-leaf stage), culms (tillering
                     stage and flowering  stage), roots, flag leaves, panicles
                     (flowering stage and two different  stages of
                     grain-filling), flowers, leaves after Xanthomonas oryzae
                     pv oryzae  (strain PXO61) inoculation, and leaves after
                     Pyricularia grisea (isolate  1814) inoculation"
ORIGIN

Alignment Scores:
Pred. No.:                2.44          Length:          643
Score:                    74.00         Matches:         10
Percent Similarity:       100.00%       Conservative:    2
Best Local Similarity:    83.33%        Mismatches:      0
Query Match:              88.10%        Indels:          0
DB:                       2             Gaps:            0

US-09-858-332G-15 (1-12) x BF108353 (1-643)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::||||||||||||||||||||||||
Db        181 CACCACCACGACCACAACCACAACCACAACCACAAC 216

RESULT 267
CZ292359/c
LOCUS       CZ292359                715 bp    DNA     linear   GSS 01-JUL-2005
DEFINITION  cp80h11.f Candida parapsilosis Random Genomic Library Candida
            parapsilosis genomic clone cp80h11, genomic survey sequence.
ACCESSION   CZ292359
VERSION     CZ292359.1  GI:68463669
KEYWORDS    GSS.
SOURCE      Candida parapsilosis
  ORGANISM  Candida parapsilosis
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 715)
  AUTHORS   Logue,M.E., Wong,S., Wolfe,K.H. and Butler,G.
  TITLE     A genome sequence survey shows that the pathogenic yeast Candida
            parapsilosis has a defective MTLa1 allele at its mating type locus
  JOURNAL   Eukaryot. Cell 4 (6), 1009-1017 (2005)
  PUBMED    15947193
COMMENT     Contact: Logue M
            Department of Biochemistry, Conway Institute of Biomolecular and
            Biomedical Research
            University College Dublin
            Dublin 4, Ireland
            Tel: +353 1 7166885
            Fax: +353 1 2837211
            Email: mary.e.logue@ucd.ie
            Class: plasmid ends.
FEATURES             Location/Qualifiers
     source          1. .715
                     /organism="Candida parapsilosis"
                     /mol_type="genomic DNA"
                     /strain="CLIB214"
                     /db_xref="taxon:5480"
                     /clone="cp80h11"
                     /clone_lib="Candida parapsilosis Random Genomic Library"
ORIGIN

Alignment Scores:
Pred. No.:                2.72          Length:          715
Score:                    74.00         Matches:         10
Percent Similarity:       100.00%       Conservative:    2
Best Local Similarity:    83.33%        Mismatches:      0
Query Match:              88.10%        Indels:          0
DB:                       10            Gaps:            0

US-09-858-332G-15 (1-12) x CZ292359 (1-715)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||||||||||||||||||||||||||:::
Db        478 CACTCTCACAACCACAACCACAACCACAACCACCAC 443

RESULT 268
BX900547
LOCUS       BX900547                792 bp    mRNA    linear   EST 16-JAN-2004
DEFINITION  BX900547 Oryza sativa library (Han B) Oryza sativa cDNA clone
            y708b01p5, mRNA sequence.
ACCESSION   BX900547
VERSION     BX900547.1  GI:40492993
KEYWORDS    EST.
SOURCE      Oryza sativa
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 792)
  AUTHORS   Han,B., Feng,Q., Huang,Y.C., Ying,K., Li,Y., Guan,J.P., Zhu,J.J.,
            Zhao,Q., Hu,X., Liu,Y.L., Mu,J., Yu,Z., Chen,L., Fan,D.L.,
            Weng,Q.J., Zhang,L., Lu,Y.Q., Yu,S.L., Liu,X.H., Lu,T.T.,
            Zhang,Y.J., Lu,Y., Li,C., Li,T., Zhang,Y., Hu,H., Jia,P.X.,
            Zhang,L., Lan,L.F., Chen,W., Wu,S.A. and Xue,Y.B.
```

```
  TITLE     Rice cDNA EST clone
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Han Bin
            National Center for Gene Research
            Chinese Academy of Sciences
            500# Cao Bao Road, Shanghai 200233, China
            Email: bhan@ncgr.ac.cn
            Clone requests: bhan@ncgr.ac.cn
             This is rice cdna est clone
            Web site: http://www.ncgr.ac.cn.
FEATURES             Location/Qualifiers
     source          1. .792
                     /organism="Oryza sativa"
                     /mol_type="mRNA"
                     /db_xref="taxon:4530"
                     /clone="y708b01p5"
                     /clone_lib="Oryza sativa library (Han B)"
ORIGIN

Alignment Scores:
Pred. No.:              3.03          Length:         792
Score:                  74.00         Matches:        10
Percent Similarity:     100.00%       Conservative:   2
Best Local Similarity:  83.33%        Mismatches:     0
Query Match:            88.10%        Indels:         0
DB:                     5             Gaps:           0

US-09-858-332G-15 (1-12) x BX900547 (1-792)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::||||||||||||||||||||||||
Db        241 CACCACCACGACCACAACCACAACCACAACCACAAC 276

RESULT 269
DR027045/c
LOCUS       DR027045                798 bp    mRNA    linear   EST 26-MAY-2005
DEFINITION  Osmo01078 F. cylindrus osmotic stress library Fragilariopsis
            cylindrus cDNA clone FcylESTa47g07.s1, mRNA sequence.
ACCESSION   DR027045
VERSION     DR027045.1  GI:66749408
KEYWORDS    EST.
SOURCE      Fragilariopsis cylindrus
  ORGANISM  Fragilariopsis cylindrus
            Eukaryota; stramenopiles; Bacillariophyta; Bacillariophyceae;
            Bacillariophycidae; Bacillariales; Bacillariaceae; Fragilariopsis.
REFERENCE   1  (bases 1 to 798)
  AUTHORS   Krell,A. and Gloeckner,G.
  TITLE     Analysis of an osmotic stress induced cDNA library of the
            psychrophilic diatom Fragilariopsis cylindrus
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Krell, Andreas; Gloeckner, Gernot
            Biological Oceanography, Sea ice research; Genome Analysis
            Alfred-Wegner-Institute for Polar and Marine Research; Institute
            for Molecular Biotechnology
            Am Handelshafen 12, D-27570 Bremerhaven, Germany; Beutenbergstr.
            11, D-07745 Jena, Germany
            Tel: ++49 471 48311812; ++49 3641 656440
            Fax: ++49 471 48311425; ++49 3641 656255
            Email: akrell@awi-bremerhaven.de; gernot@imb-jena.de
            PCR PRimers
            FORWARD: 5'M13
            BACKWARD: 3'M13
            Seq primer: 5'GTAAAACGACGGCCAG 3'.
FEATURES             Location/Qualifiers
     source          1. .798
                     /organism="Fragilariopsis cylindrus"
                     /mol_type="mRNA"
                     /db_xref="taxon:186039"
                     /clone="FcylESTa47g07.s1"
                     /clone_lib="F. cylindrus osmotic stress library"
                     /note="Samples for total RNA isolation were taken
                     continous for 5 days after a salt shock treatment
                     increasing salinity from 34 to 60 PSU. Total RNA
                     extraktion was performed with RNAqueous (Ambion) and mRNA
                     purification with polyA Purist (Ambion). Further steps
                     were carried out as described in the Cloneminer Kit. cDNA
                     size fractionation was carried out with CHROMA Spin-400
                     columns and additionally on a gel."
ORIGIN

Alignment Scores:
Pred. No.:              3.06          Length:         798
Score:                  74.00         Matches:        10
Percent Similarity:     100.00%       Conservative:   2
Best Local Similarity:  83.33%        Mismatches:     0
Query Match:            88.10%        Indels:         0
DB:                     8             Gaps:           0

US-09-858-332G-15 (1-12) x DR027045 (1-798)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||||||||:::||||||||||||||:::||||||
Db        794 CATAACCACCACCATAACCATAACCACCACCATAAC 759

RESULT 270
CK251091
LOCUS       CK251091                803 bp    mRNA    linear   EST 30-JUL-2004
DEFINITION  EST734728 potato callus cDNA library, normalized and full-length
            Solanum tuberosum cDNA clone POCBK84 5' end, mRNA sequence.
ACCESSION   CK251091
VERSION     CK251091.1  GI:39803793
KEYWORDS    EST.
SOURCE      Solanum tuberosum (potato)
  ORGANISM  Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum.
REFERENCE   1  (bases 1 to 803)
  AUTHORS   Buell,C.R., Hart,A., Zismann,V., Karamycheva,S.A. and Baker,B.
  TITLE     Generation of ESTs from potato callus tissue
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Robin Buell
            The Institute for Genomic Research
            9712 Medical Center Dr, Rockville, MD 20850, USA
            Email: potato-array@tigr.org
            Clones can be requested from the University of Arizona Genomics
            Institute via http://genome.arizona.edu/orders/ .
            Seq primer: ATT TAG GTG ACA CTA TAG.
FEATURES             Location/Qualifiers
     source          1. .803
                     /organism="Solanum tuberosum"
                     /mol_type="mRNA"
                     /cultivar="Kennebec"
                     /db_xref="taxon:4113"
                     /clone="POCBK84"
                     /tissue_type="callus"
                     /lab_host="DH10B-TonA"
                     /clone_lib="potato callus cDNA library, normalized and
                     full-length"
                     /note="Vector: pCMVSport6.1; Site_1: EcoRI; Site_2: NotI;
                     supplier: RNA was isolated from Solanum tuberosum var.
                     Kennebec callus tissue grown on solid media."
ORIGIN

Alignment Scores:
Pred. No.:              3.08          Length:         803
Score:                  74.00         Matches:        10
Percent Similarity:     100.00%       Conservative:   2
Best Local Similarity:  83.33%        Mismatches:     0
Query Match:            88.10%        Indels:         0
DB:                     7             Gaps:           0

US-09-858-332G-15 (1-12) x CK251091 (1-803)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||:::|||:::||||||
```

```
Db        663 CACAACCACAACCACAACCACGATCACGACCACAAC 698

RESULT 271
CX114190
LOCUS       CX114190                804 bp    mRNA    linear   EST 03-JUN-2005
DEFINITION  EI086F05 A normalized whole-life-cycle cDNA library of rice Oryza
            sativa (indica cultivar-group) cDNA clone EI086F05 5', mRNA
            sequence.
ACCESSION   CX114190
VERSION     CX114190.1  GI:66927342
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 804)
  AUTHORS   Zhang,J., Feng,Q., Jin,C., Qiu,D., Zhang,L., Xie,K., Yuan,D.,
            Han,B., Zhang,Q. and Wang,S.
  TITLE     Features of the expressed sequences revealed by a large-scale
            analysis of ESTs from a normalized cDNA library of the elite indica
            rice cultivar Minghui 63
  JOURNAL   Plant J. 42 (5), 772-780 (2005)
   PUBMED   15918889
COMMENT     Contact: Wang S
            National Key Laboratory of Crop Genetic Improvement
            Hauzhong Agricultural University
            Wuhan 430070, China
            Tel: 86-27-87282044
            Fax: 86-27-87287092
            Email: shipingwang@hotmail.com
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .804
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /strain="indica"
                     /cultivar="Minghui 63"
                     /db_xref="taxon:39946"
                     /clone="EI086F05"
                     /tissue_type="whole plant"
                     /dev_stage="whole-life-cycle"
                     /lab_host="E. coli DH10B"
                     /clone_lib="A normalized whole-life-cycle cDNA library of
                     rice"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; The
                     library is constructed based on the strategy of satuation
                     hybridization with genomic DNA using rice cultivar Minghui
                     63. This library consists of cDNA from 15 directionally
                     cloned cDNA libraries constructed with different tissues
                     from 9 developmental stages."
ORIGIN

Alignment Scores:
Pred. No.:                3.08          Length:          804
Score:                    74.00         Matches:         10
Percent Similarity:       100.00%       Conservative:    2
Best Local Similarity:    83.33%        Mismatches:      0
Query Match:              88.10%        Indels:          0
DB:                       8             Gaps:            0

US-09-858-332G-15 (1-12) x CX114190 (1-804)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        218 CACCACCACGACCACAACCACAACCACAACCACAAC 253

RESULT 272
CX669339/c
LOCUS       CX669339                813 bp    mRNA    linear   EST 19-JAN-2005
DEFINITION  UCRCP01_044_C12_T7 Swingle citrumelo nematode-challenged root cDNA
            library - UCRCP01 Citrus x paradisi x Poncirus trifoliata cDNA
            clone UCRCP01_044_T7_C12, mRNA sequence.
ACCESSION   CX669339
VERSION     CX669339.1  GI:57928288
KEYWORDS    EST.
SOURCE      Citrus x paradisi x Poncirus trifoliata
  ORGANISM  Citrus x paradisi x Poncirus trifoliata
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Sapindales; Rutaceae; Citrus x Poncirus.
REFERENCE   1  (bases 1 to 813)
  AUTHORS   Close,T.J., Roose,M.L., Becker,O., Darsow,J., Federici,C.F.,
            Fenton,R.D., Wanamaker,S., Choi,Y.G and Kingan,T.
  TITLE     Development of EST Resources and New Genetic Markers for California
            Citrus - Swingle citrumelo nematode-challenged roots - UCRCP01
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Timothy J. Close
            Department of Botany & Plant Sciences
            University of California
            Riverside, CA 92521-0124, USA
            Tel: 909-787-3318
            Fax: 909-787-4437
            Email: timothy.close@ucr.edu
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .813
                     /organism="Citrus x paradisi x Poncirus trifoliata"
                     /mol_type="mRNA"
                     /cultivar="Swingle"
                     /db_xref="taxon:309804"
                     /clone="UCRCP01_044_T7_C12"
                     /tissue_type="root"
                     /dev_stage="seedling"
                     /lab_host="E. coli TJC121"
                     /clone_lib="Swingle citrumelo nematode-challenged root
                     cDNA library - UCRCP01"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; Twelve seedlings were
                     transplanted from cones to 4 inch pots of sand mix, then
                     grown in greenhouses (Darsow, Becker lab). Citrus
                     nematodes (Tylenchulus semipenitrans) were extracted from
                     contaminated soil from the UC Riverside Citrus Research
                     Center then suspended in water and added to pots (Darsow).
                     Growing roots were collected by clipping them away from
                     the old root ball 24 hr and 4 days after inoculation
                     (Federici). Total RNA was extracted using Concert Plant
                     RNA Reagent (Invitrogen)followed by RNeasy (Qiagen), an
                     equal amount of RNA was pooled from each of the two
                     collection times, poly(A) RNA was purified using Oligotex
                     (Qiagen), a cDNA library was made using a cDNA synthesis
                     kit (Stratagene), then 0.45 million primary lambda cDNA
                     clones were in vivo excised to give a population of
                     pBluescript SK(-) phagemids (Fenton, Close lab). Phagemids
                     were plated, plasmid DNA purified, cDNA clones archived,
                     and DNA sequences determined bi-directionally using an
                     ABI3730 at the University of California Riverside
                     Institute of Integrative Genome Biology Genomics Core
                     Instrumentation Facility, (Choi, Kingan). Chromatogram
                     files were downloaded by FTP by Close, then processed by
                     Wanamaker (Close lab) using the HarvEST pipeline
                     (http://harvest.ucr.edu) to remove vector and cloning
                     oligo sequences and various contaminants, and to trim to a
                     high quality region. Sequences that retained a phred 17
                     region of at least 100 bases were assembled, then chimeras
                     were removed following manual inspection of assemblies
                     (Close, Roose, Wanamaker). Sequences that survived all
                     removal steps were submitted to GenBank."
ORIGIN

Alignment Scores:
Pred. No.:                3.12          Length:          813
Score:                    74.00         Matches:         10
Percent Similarity:       100.00%       Conservative:    2
Best Local Similarity:    83.33%        Mismatches:      0
Query Match:              88.10%        Indels:          0
```

```
DB:                    8             Gaps:          0

US-09-858-332G-15 (1-12) x CX669339 (1-813)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||:::|||:::||||||
Db        563 CACAATCACAATCACAATCACAGTCACAGTCACAAT 528

RESULT 273
CX118645
LOCUS       CX118645                818 bp    mRNA    linear   EST 03-JUN-2005
DEFINITION  EI116B05 A normalized whole-life-cycle cDNA library of rice Oryza
            sativa (indica cultivar-group) cDNA clone EI116B05 5', mRNA
            sequence.
ACCESSION   CX118645
VERSION     CX118645.1  GI:66931797
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 818)
  AUTHORS   Zhang,J., Feng,Q., Jin,C., Qiu,D., Zhang,L., Xie,K., Yuan,D.,
            Han,B., Zhang,Q. and Wang,S.
  TITLE     Features of the expressed sequences revealed by a large-scale
            analysis of ESTs from a normalized cDNA library of the elite indica
            rice cultivar Minghui 63
  JOURNAL   Plant J. 42 (5), 772-780 (2005)
   PUBMED   15918889
COMMENT     Contact: Wang S
            National Key Laboratory of Crop Genetic Improvement
            Hauzhong Agricultural University
            Wuhan 430070, China
            Tel: 86-27-87282044
            Fax: 86-27-87287092
            Email: shipingwang@hotmail.com
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .818
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /strain="indica"
                     /cultivar="Minghui 63"
                     /db_xref="taxon:39946"
                     /clone="EI116B05"
                     /tissue_type="whole plant"
                     /dev_stage="whole-life-cycle"
                     /lab_host="E. coli DH10B"
                     /clone_lib="A normalized whole-life-cycle cDNA library of
                     rice"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; The
                     library is constructed based on the strategy of satuation
                     hybridization with genomic DNA using rice cultivar Minghui
                     63. This library consists of cDNA from 15 directionally
                     cloned cDNA libraries constructed with different tissues
                     from 9 developmental stages."
ORIGIN

Alignment Scores:
Pred. No.:              3.14          Length:        818
Score:                  74.00         Matches:       10
Percent Similarity:     100.00%       Conservative:  2
Best Local Similarity:  83.33%        Mismatches:    0
Query Match:            88.10%        Indels:        0
DB:                     8             Gaps:          0

US-09-858-332G-15 (1-12) x CX118645 (1-818)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::||||||||||||||||||||||||
Db        220 CACCACCACGACCACAACCACAACCACAACCACAAC 255

RESULT 274
CZ287109/c
LOCUS       CZ287109                854 bp    DNA     linear   GSS 01-JUL-2005
DEFINITION  cp50f12.f Candida parapsilosis Random Genomic Library Candida
            parapsilosis genomic clone cp50f12, genomic survey sequence.
ACCESSION   CZ287109
VERSION     CZ287109.1  GI:68458419
KEYWORDS    GSS.
SOURCE      Candida parapsilosis
  ORGANISM  Candida parapsilosis
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 854)
  AUTHORS   Logue,M.E., Wong,S., Wolfe,K.H. and Butler,G.
  TITLE     A genome sequence survey shows that the pathogenic yeast Candida
            parapsilosis has a defective MTLa1 allele at its mating type locus
  JOURNAL   Eukaryot. Cell 4 (6), 1009-1017 (2005)
   PUBMED   15947193
COMMENT     Contact: Logue M
            Department of Biochemistry, Conway Institute of Biomolecular and
            Biomedical Research
            University College Dublin
            Dublin 4, Ireland
            Tel: +353 1 7166885
            Fax: +353 1 2837211
            Email: mary.e.logue@ucd.ie
            Class: plasmid ends.
FEATURES             Location/Qualifiers
     source          1. .854
                     /organism="Candida parapsilosis"
                     /mol_type="genomic DNA"
                     /strain="CLIB214"
                     /db_xref="taxon:5480"
                     /clone="cp50f12"
                     /clone_lib="Candida parapsilosis Random Genomic Library"
ORIGIN

Alignment Scores:
Pred. No.:              3.28          Length:        854
Score:                  74.00         Matches:       10
Percent Similarity:     100.00%       Conservative:  2
Best Local Similarity:  83.33%        Mismatches:    0
Query Match:            88.10%        Indels:        0
DB:                     10            Gaps:          0

US-09-858-332G-15 (1-12) x CZ287109 (1-854)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::||||||||||||||||||||||||||:::
Db        460 CACTCTCACAACCACAACCACAACCACAACCACCAC 425

RESULT 275
CZ292089/c
LOCUS       CZ292089                862 bp    DNA     linear   GSS 01-JUL-2005
DEFINITION  cp79d09.r Candida parapsilosis Random Genomic Library Candida
            parapsilosis genomic clone cp79d09, genomic survey sequence.
ACCESSION   CZ292089
VERSION     CZ292089.1  GI:68463399
KEYWORDS    GSS.
SOURCE      Candida parapsilosis
  ORGANISM  Candida parapsilosis
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 862)
  AUTHORS   Logue,M.E., Wong,S., Wolfe,K.H. and Butler,G.
  TITLE     A genome sequence survey shows that the pathogenic yeast Candida
            parapsilosis has a defective MTLa1 allele at its mating type locus
  JOURNAL   Eukaryot. Cell 4 (6), 1009-1017 (2005)
   PUBMED   15947193
COMMENT     Contact: Logue M
            Department of Biochemistry, Conway Institute of Biomolecular and
            Biomedical Research
            University College Dublin
```

Dublin 4, Ireland
Tel: +353 1 7166885
Fax: +353 1 2837211
Email: mary.e.logue@ucd.ie
Class: plasmid ends.

FEATURES Location/Qualifiers
source 1. .862
/organism="Candida parapsilosis"
/mol_type="genomic DNA"
/strain="CLIB214"
/db_xref="taxon:5480"
/clone="cp79d09"
/clone_lib="Candida parapsilosis Random Genomic Library"

ORIGIN

Alignment Scores:

Pred. No.:	3.31	Length:	862
Score:	74.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	83.33%	Mismatches:	0
Query Match:	88.10%	Indels:	0
DB:	10	Gaps:	0

US-09-858-332G-15 (1-12) x CZ292089 (1-862)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::||||||||||||||||||||||||||:::
Db        460 CACTCTCACAACCACAACCACAACCACAACCACCAC 425
```

RESULT 276
CX669338
LOCUS CX669338 876 bp mRNA linear EST 19-JAN-2005
DEFINITION UCRCP01_044_C12_T3 Swingle citrumelo nematode-challenged root cDNA library - UCRCP01 Citrus x paradisi x Poncirus trifoliata cDNA clone UCRCP01_044_T3_C12, mRNA sequence.
ACCESSION CX669338
VERSION CX669338.1 GI:57928287
KEYWORDS EST.
SOURCE Citrus x paradisi x Poncirus trifoliata
ORGANISM Citrus x paradisi x Poncirus trifoliata
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Sapindales; Rutaceae; Citrus x Poncirus.
REFERENCE 1 (bases 1 to 876)
AUTHORS Close,T.J., Roose,M.L., Becker,O., Darsow,J., Federici,C.F., Fenton,R.D., Wanamaker,S., Choi,Y.G and Kingan,T.
TITLE Development of EST Resources and New Genetic Markers for California Citrus - Swingle citrumelo nematode-challenged roots - UCRCP01
JOURNAL Unpublished (2005)
COMMENT Contact: Timothy J. Close
Department of Botany & Plant Sciences
University of California
Riverside, CA 92521-0124, USA
Tel: 909-787-3318
Fax: 909-787-4437
Email: timothy.close@ucr.edu
Seq primer: T3.

FEATURES Location/Qualifiers
source 1. .876
/organism="Citrus x paradisi x Poncirus trifoliata"
/mol_type="mRNA"
/cultivar="Swingle"
/db_xref="taxon:309804"
/clone="UCRCP01_044_T3_C12"
/tissue_type="root"
/dev_stage="seedling"
/lab_host="E. coli TJC121"
/clone_lib="Swingle citrumelo nematode-challenged root cDNA library - UCRCP01"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site_1: EcoRI; Site_2: XhoI; Twelve seedlings were transplanted from cones to 4 inch pots of sand mix, then grown in greenhouses (Darsow, Becker lab). Citrus nematodes (Tylenchulus semipenitrans) were extracted from contaminated soil from the UC Riverside Citrus Research Center then suspended in water and added to pots (Darsow). Growing roots were collected by clipping them away from the old root ball 24 hr and 4 days after inoculation (Federici). Total RNA was extracted using Concert Plant RNA Reagent (Invitrogen) followed by RNeasy (Qiagen), an equal amount of RNA was pooled from each of the two collection times, poly(A) RNA was purified using Oligotex (Qiagen), a cDNA library was made using a cDNA synthesis kit (Stratagene), then 0.45 million primary lambda cDNA clones were in vivo excised to give a population of pBluescript SK(-) phagemids (Fenton, Close lab). Phagemids were plated, plasmid DNA purified, cDNA clones archived, and DNA sequences determined bi-directionally using an ABI3730 at the University of California Riverside Institute of Integrative Genome Biology Genomics Core Instrumentation Facility, (Choi, Kingan). Chromatogram files were downloaded by FTP by Close, then processed by Wanamaker (Close lab) using the HarvEST pipeline (http://harvest.ucr.edu) to remove vector and cloning oligo sequences and various contaminants, and to trim to a high quality region. Sequences that retained a phred 17 region of at least 100 bases were assembled, then chimeras were removed following manual inspection of assemblies (Close, Roose, Wanamaker). Sequences that survived all removal steps were submitted to GenBank."

ORIGIN

Alignment Scores:

Pred. No.:	3.37	Length:	876
Score:	74.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	83.33%	Mismatches:	0
Query Match:	88.10%	Indels:	0
DB:	8	Gaps:	0

US-09-858-332G-15 (1-12) x CX669338 (1-876)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||:::|||:::||||||
Db        525 CACAATCACAATCACAATCACAGTCACAGTCACAAT 560
```

RESULT 277
CX104704
LOCUS CX104704 899 bp mRNA linear EST 03-JUN-2005
DEFINITION BI104A04 A normalized whole-life-cycle cDNA library of rice Oryza sativa (indica cultivar-group) cDNA clone BI104A04 5', mRNA sequence.
ACCESSION CX104704
VERSION CX104704.1 GI:66917856
KEYWORDS EST.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 899)
AUTHORS Zhang,J., Feng,Q., Jin,C., Qiu,D., Zhang,L., Xie,K., Yuan,D., Han,B., Zhang,Q. and Wang,S.
TITLE Features of the expressed sequences revealed by a large-scale analysis of ESTs from a normalized cDNA library of the elite indica rice cultivar Minghui 63
JOURNAL Plant J. 42 (5), 772-780 (2005)
PUBMED 15918889
COMMENT Contact: Wang S
National Key Laboratory of Crop Genetic Improvement
Hauzhong Agricultural University
Wuhan 430070, China
Tel: 86-27-87282044
Fax: 86-27-87287092
Email: shipingwang@hotmail.com
Seq primer: T7.

```
FEATURES             Location/Qualifiers
     source          1. .899
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /strain="indica"
                     /cultivar="Minghui 63"
                     /db_xref="taxon:39946"
                     /clone="BI104A04"
                     /tissue_type="whole plant"
                     /dev_stage="whole-life-cycle"
                     /lab_host="E. coli DH10B"
                     /clone_lib="A normalized whole-life-cycle cDNA library of
                     rice"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; The
                     library is constructed based on the strategy of satuation
                     hybridization with genomic DNA using rice cultivar Minghui
                     63. This library consists of cDNA from 15 directionally
                     cloned cDNA libraries constructed with different tissues
                     from 9 developmental stages."
ORIGIN

Alignment Scores:
Pred. No.:              3.46          Length:         899
Score:                  74.00         Matches:        10
Percent Similarity:     100.00%       Conservative:   2
Best Local Similarity:  83.33%        Mismatches:     0
Query Match:            88.10%        Indels:         0
DB:                     8             Gaps:           0

US-09-858-332G-15 (1-12) x CX104704 (1-899)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        228 CACCACCACGACCACAACCACAACCACAACCACAAC 263

RESULT 278
CX112129
LOCUS       CX112129                899 bp    mRNA    linear   EST 03-JUN-2005
DEFINITION  EI074E01 A normalized whole-life-cycle cDNA library of rice Oryza
            sativa (indica cultivar-group) cDNA clone EI074E01 5', mRNA
            sequence.
ACCESSION   CX112129
VERSION     CX112129.1  GI:66925281
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 899)
  AUTHORS   Zhang,J., Feng,Q., Jin,C., Qiu,D., Zhang,L., Xie,K., Yuan,D.,
            Han,B., Zhang,Q. and Wang,S.
  TITLE     Features of the expressed sequences revealed by a large-scale
            analysis of ESTs from a normalized cDNA library of the elite indica
            rice cultivar Minghui 63
  JOURNAL   Plant J. 42 (5), 772-780 (2005)
   PUBMED   15918889
COMMENT     Contact: Wang S
            National Key Laboratory of Crop Genetic Improvement
            Hauzhong Agricultural University
            Wuhan 430070, China
            Tel: 86-27-87282044
            Fax: 86-27-87287092
            Email: shipingwang@hotmail.com
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .899
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /strain="indica"
                     /cultivar="Minghui 63"
                     /db_xref="taxon:39946"
                     /clone="EI074E01"
                     /tissue_type="whole plant"
                     /dev_stage="whole-life-cycle"
                     /lab_host="E. coli DH10B"
                     /clone_lib="A normalized whole-life-cycle cDNA library of
                     rice"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; The
                     library is constructed based on the strategy of satuation
                     hybridization with genomic DNA using rice cultivar Minghui
                     63. This library consists of cDNA from 15 directionally
                     cloned cDNA libraries constructed with different tissues
                     from 9 developmental stages."
ORIGIN

Alignment Scores:
Pred. No.:              3.46          Length:         899
Score:                  74.00         Matches:        10
Percent Similarity:     100.00%       Conservative:   2
Best Local Similarity:  83.33%        Mismatches:     0
Query Match:            88.10%        Indels:         0
DB:                     8             Gaps:           0

US-09-858-332G-15 (1-12) x CX112129 (1-899)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        228 CACCACCACGACCACAACCACAACCACAACCACAAC 263

RESULT 279
DU040292
LOCUS       DU040292                932 bp    DNA     linear   GSS 12-AUG-2005
DEFINITION  25505 Tomato HindIII BAC Library Lycopersicon esculentum genomic
            clone LE_HBa0089L09 5, genomic survey sequence.
ACCESSION   DU040292
VERSION     DU040292.1  GI:72461346
KEYWORDS    GSS.
SOURCE      Lycopersicon esculentum (Solanum lycopersicum)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 932)
  AUTHORS   Mueller,L.A., Buels,R.M., Wang,Y., Tanksley,S.D., Giovannoni,J.J.,
            Van Eck,J. and Stack,S.
  TITLE     BAC end sequencing from three Solanum lycopersicon libraries
  JOURNAL   Unpublished (2005)
COMMENT     Other_GSSs: 25504
            Contact: Lukas Mueller
            Tanksley Lab, Dept. of Plant Breeding
            Cornell University
            251 Emerson Hall, Ithaca, NY 14853, USA
            Tel: 607-255-6557
            Fax: 607-255-6683
            Email: sgn-feedback@sgn.cornell.edu
            Plate: 89  row: L  column: 9
            Seq primer: T7
            Class: BAC ends
            High quality sequence start: 48
            High quality sequence stop: 716.
FEATURES             Location/Qualifiers
     source          1. .932
                     /organism="Lycopersicon esculentum"
                     /mol_type="genomic DNA"
                     /cultivar="Heinz 1706"
                     /db_xref="taxon:4081"
                     /clone="LE_HBa0089L09"
                     /lab_host="E. coli"
                     /clone_lib="Tomato HindIII BAC Library"
                     /note="Vector: pBeloBAC11; Site_1: HindIII"
ORIGIN

Alignment Scores:
Pred. No.:              3.59          Length:         932
Score:                  74.00         Matches:        10
```

```
Percent Similarity:     100.00%         Conservative:  2
Best Local Similarity:  83.33%          Mismatches:    0
Query Match:            88.10%          Indels:        0
DB:                     10              Gaps:          0

US-09-858-332G-15 (1-12) x DU040292 (1-932)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||:::|||:::||||||
Db        457 CACAACCACAACCACAACCACCACCACCACCACAAC 492

RESULT 280
CX118565
LOCUS       CX118565                942 bp    mRNA    linear   EST 03-JUN-2005
DEFINITION  EI115L11 A normalized whole-life-cycle cDNA library of rice Oryza
            sativa (indica cultivar-group) cDNA clone EI115L11 5', mRNA
            sequence.
ACCESSION   CX118565
VERSION     CX118565.1  GI:66931717
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 942)
  AUTHORS   Zhang,J., Feng,Q., Jin,C., Qiu,D., Zhang,L., Xie,K., Yuan,D.,
            Han,B., Zhang,Q. and Wang,S.
  TITLE     Features of the expressed sequences revealed by a large-scale
            analysis of ESTs from a normalized cDNA library of the elite indica
            rice cultivar Minghui 63
  JOURNAL   Plant J. 42 (5), 772-780 (2005)
   PUBMED   15918889
COMMENT     Contact: Wang S
            National Key Laboratory of Crop Genetic Improvement
            Hauzhong Agricultural University
            Wuhan 430070, China
            Tel: 86-27-87282044
            Fax: 86-27-87287092
            Email: shipingwang@hotmail.com
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .942
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /strain="indica"
                     /cultivar="Minghui 63"
                     /db_xref="taxon:39946"
                     /clone="EI115L11"
                     /tissue_type="whole plant"
                     /dev_stage="whole-life-cycle"
                     /lab_host="E. coli DH10B"
                     /clone_lib="A normalized whole-life-cycle cDNA library of
                     rice"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; The
                     library is constructed based on the strategy of satuation
                     hybridization with genomic DNA using rice cultivar Minghui
                     63. This library consists of cDNA from 15 directionally
                     cloned cDNA libraries constructed with different tissues
                     from 9 developmental stages."
ORIGIN

Alignment Scores:
Pred. No.:              3.63            Length:        942
Score:                  74.00           Matches:       10
Percent Similarity:     100.00%         Conservative:  2
Best Local Similarity:  83.33%          Mismatches:    0
Query Match:            88.10%          Indels:        0
DB:                     8               Gaps:          0

US-09-858-332G-15 (1-12) x CX118565 (1-942)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::||||||||||||||||||||||||
Db        237 CACCACCACGACCACAACCACAACCACAACCACAAC 272

RESULT 281
CNS07CAB
LOCUS       CNS07CAB                989 bp    DNA     linear   GSS 08-JUL-2001
DEFINITION  T3 end of clone BD0AA001C03 of library BD0AA from strain CBS 94 of
            Candida tropicalis, genomic survey sequence.
ACCESSION   AL438921
VERSION     AL438921.1  GI:12222334
KEYWORDS    GSS.
SOURCE      Candida tropicalis
  ORGANISM  Candida tropicalis
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 989)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
   PUBMED   11152876
REFERENCE   2  (bases 1 to 989)
  AUTHORS   Blandin,G., Ozier-Kalogeropoulos,O., Wincker,P., Artiguenave,F. and
            Dujon,B.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 16. Candida
            tropicalis
  JOURNAL   FEBS Lett. 487 (1), 91-94 (2000)
   PUBMED   11152891
REFERENCE   3  (bases 1 to 989)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web :
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .989
                     /organism="Candida tropicalis"
                     /mol_type="genomic DNA"
                     /strain="CBS 94"
                     /db_xref="taxon:5482"
                     /clone="BD0AA001C03"
                     /clone_lib="BD0AA"
                     /note="end : T3"
ORIGIN

Alignment Scores:
Pred. No.:              3.82            Length:        989
Score:                  74.00           Matches:       10
Percent Similarity:     100.00%         Conservative:  2
Best Local Similarity:  83.33%          Mismatches:    0
Query Match:            88.10%          Indels:        0
DB:                     11              Gaps:          0

US-09-858-332G-15 (1-12) x CNS07CAB (1-989)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||:::|||:::||||||||||||
Db        636 CACAATCACAATCACCACCACGATCACAATCACAAC 671
```

```
RESULT 282
AG612034/c
LOCUS       AG612034               1032 bp    DNA     linear   GSS 23-DEC-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-547H03.TJ, genomic survey
            sequence.
ACCESSION   AG612034
VERSION     AG612034.1  GI:48372864
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus (Japanese wild mouse)
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
            Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
            Shiroishi,T.
  TITLE     Contribution of Asian mouse subspecies Mus musculus molossinus to
            genomic constitution of strain C57BL/6J, as defined by BAC-end
            sequence-SNP analysis
  JOURNAL   Genome Res. 14 (12), 2439-2447 (2004)
   PUBMED   15574823
REFERENCE   2  (bases 1 to 1032)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1032
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-547H03.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

Alignment Scores:
Pred. No.:              4              Length:        1032
Score:                  74.00          Matches:       10
Percent Similarity:     100.00%        Conservative:  2
Best Local Similarity:  83.33%         Mismatches:    0
Query Match:            88.10%         Indels:        0
DB:                     10             Gaps:          0

US-09-858-332G-15 (1-12) x AG612034 (1-1032)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        198 CATCACCACCACCACAACCACAATCACAACCACAAC 163

RESULT 283
CX115929
LOCUS       CX115929               1076 bp    mRNA    linear   EST 03-JUN-2005
DEFINITION  EI101J02 A normalized whole-life-cycle cDNA library of rice Oryza
            sativa (indica cultivar-group) cDNA clone EI101J02 5', mRNA
            sequence.
ACCESSION   CX115929
VERSION     CX115929.1  GI:66929081
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 1076)
  AUTHORS   Zhang,J., Feng,Q., Jin,C., Qiu,D., Zhang,L., Xie,K., Yuan,D.,
            Han,B., Zhang,Q. and Wang,S.
  TITLE     Features of the expressed sequences revealed by a large-scale
            analysis of ESTs from a normalized cDNA library of the elite indica
            rice cultivar Minghui 63
  JOURNAL   Plant J. 42 (5), 772-780 (2005)
   PUBMED   15918889
COMMENT     Contact: Wang S
            National Key Laboratory of Crop Genetic Improvement
            Hauzhong Agricultural University
            Wuhan 430070, China
            Tel: 86-27-87282044
            Fax: 86-27-87287092
            Email: shipingwang@hotmail.com
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .1076
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /strain="indica"
                     /cultivar="Minghui 63"
                     /db_xref="taxon:39946"
                     /clone="EI101J02"
                     /tissue_type="whole plant"
                     /dev_stage="whole-life-cycle"
                     /lab_host="E. coli DH10B"
                     /clone_lib="A normalized whole-life-cycle cDNA library of
                     rice"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; The
                     library is constructed based on the strategy of satuation
                     hybridization with genomic DNA using rice cultivar Minghui
                     63. This library consists of cDNA from 15 directionally
                     cloned cDNA libraries constructed with different tissues
                     from 9 developmental stages."
ORIGIN

Alignment Scores:
Pred. No.:              4.18           Length:        1076
Score:                  74.00          Matches:       10
Percent Similarity:     100.00%        Conservative:  2
Best Local Similarity:  83.33%         Mismatches:    0
Query Match:            88.10%         Indels:        0
DB:                     8              Gaps:          0

US-09-858-332G-15 (1-12) x CX115929 (1-1076)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        237 CACCACCACGACCACAACCACAACCACAACCACAAC 272

RESULT 284
BE035458
LOCUS       BE035458               1171 bp    mRNA    linear   EST 07-JUN-2000
DEFINITION  MO06B10 MO Mesembryanthemum crystallinum cDNA 5', mRNA sequence.
ACCESSION   BE035458
VERSION     BE035458.1  GI:8330467
KEYWORDS    EST.
SOURCE      Mesembryanthemum crystallinum (common iceplant)
  ORGANISM  Mesembryanthemum crystallinum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
```

Caryophyllales; Aizoaceae; Mesembryanthemum.
REFERENCE 1 (bases 1 to 1171)
AUTHORS Bohnert,H.J., Borchert,C., Brazille,S., Brooks,J., Eaton,M., Ferrea,H., Kawasaki,S., McCollough,A., Michalowski,C.B., Palacio,C., Scara,G., Wheeler,M. and Zepeda,G.R.
TITLE Functional Genomics of Plant Stress Tolerance
JOURNAL Unpublished (2000)
COMMENT Contact: Michalowski,C.B.
University of Arizona
Bio Sciences West room 513, Tucson, AZ 85721, USA
Tel: 520-621-7982
Fax: 520-621-1697
Email: cbm@u.arizona.edu
Best blastx match: 'gi|4056425 (AC005322) ESTs gb|H36249: gb|AA59732 and gb|AA651219. . . 135 4e-55'. An open reading frame exists.
Insert Length: 1 Std Error: 0.00.
FEATURES Location/Qualifiers
source 1. .1171
/organism="Mesembryanthemum crystallinum"
/mol_type="mRNA"
/db_xref="taxon:3544"
/tissue_type="apical meristem and leaf primordia"
/dev_stage="5 weeks"
/clone_lib="MO"
/note="no stress"
ORIGIN

Alignment Scores:

Pred. No.:	4.56	Length:	1171
Score:	74.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	83.33%	Mismatches:	0
Query Match:	88.10%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-15 (1-12) x BE035458 (1-1171)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||||||||||||||||||||||||||:::
Db        186 CATAGCCATAACCATAACCATAACCATAACCACCAC 221
```

RESULT 285
CX109102
LOCUS CX109102 1314 bp mRNA linear EST 03-JUN-2005
DEFINITION RECm0303 A normalized whole-life-cycle cDNA library of rice Oryza sativa (indica cultivar-group) cDNA clone EI008H17, EI048O20, EI047F14, FLAGLG01, EI050D16, BI030C21, EI05 5', mRNA sequence.
ACCESSION CX109102
VERSION CX109102.1 GI:66922254
KEYWORDS EST.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 1314)
AUTHORS Zhang,J., Feng,Q., Jin,C., Qiu,D., Zhang,L., Xie,K., Yuan,D., Han,B., Zhang,Q. and Wang,S.
TITLE Features of the expressed sequences revealed by a large-scale analysis of ESTs from a normalized cDNA library of the elite indica rice cultivar Minghui 63
JOURNAL Plant J. 42 (5), 772-780 (2005)
PUBMED 15918889
COMMENT Contact: Wang S
National Key Laboratory of Crop Genetic Improvement
Hauzhong Agricultural University
Wuhan 430070, China
Tel: 86-27-87282044
Fax: 86-27-87287092
Email: shipingwang@hotmail.com
Seq primer: T7.
FEATURES Location/Qualifiers
source 1. .1314
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/strain="indica"
/cultivar="Minghui 63"
/db_xref="taxon:39946"
/clone="EI008H17, EI048O20, EI047F14, FLAGLG01, EI050D16, BI030C21, EI05"
/tissue_type="whole plant"
/dev_stage="whole-life-cycle"
/lab_host="E. coli DH10B"
/clone_lib="A normalized whole-life-cycle cDNA library of rice"
/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; The library is constructed based on the strategy of satuation hybridization with genomic DNA using rice cultivar Minghui 63. This library consists of cDNA from 15 directionally cloned cDNA libraries constructed with different tissues from 9 developmental stages."
ORIGIN

Alignment Scores:

Pred. No.:	5.15	Length:	1314
Score:	74.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	83.33%	Mismatches:	0
Query Match:	88.10%	Indels:	0
DB:	8	Gaps:	0

US-09-858-332G-15 (1-12) x CX109102 (1-1314)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db        402 CACCACCACGACCACAACCACAACCACAACCACAAC 437
```

RESULT 286
AU040007
LOCUS AU040007 395 bp mRNA linear EST 29-MAR-1999
DEFINITION AU040007 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium discoideum cDNA clone SLH428, mRNA sequence.
ACCESSION AU040007
VERSION AU040007.1 GI:4009247
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 395)
AUTHORS Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M., Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H., Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
TITLE The Dictyostelium developmental cDNA project: generation and analysis of expressed sequence tags from the first-finger stage of development
JOURNAL DNA Res. 5 (6), 335-340 (1998)
PUBMED 10048482
COMMENT Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
Tel: 81-298-53-4664
Fax: 81-298-53-6614
Email: hideko@biol.tsukuba.ac.jp
PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES Location/Qualifiers
source 1. .395
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="SLH428"
/dev_stage="slug"
/clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
ORIGIN

```
Alignment Scores:
Pred. No.:                1.7           Length:        395
Score:                    73.50         Matches:       12
Percent Similarity:       92.31%        Conservative:  0
Best Local Similarity:    92.31%        Mismatches:    0
Query Match:              87.50%        Indels:        1
DB:                       1             Gaps:          1

US-09-858-332G-15 (1-12) x AU040007 (1-395)

Qy          1 HisAsnHisAsnHisAsn---HisAsnHisAsnHisAsn 12
              ||||||||||||||||||   ||||||||||||||||||
Db        267 CATAATCATAATCATAATCATCACAATCACAATCATAAT 305

RESULT 287
AU034181
LOCUS       AU034181                718 bp    mRNA    linear   EST 28-APR-1999
DEFINITION  AU034181 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
            discoideum cDNA clone SLC229, mRNA sequence.
ACCESSION   AU034181
VERSION     AU034181.1  GI:3799605
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 718)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     The Dictyostelium developmental cDNA project: generation and
            analysis of expressed sequence tags from the first-finger stage of
            development
  JOURNAL   DNA Res. 5 (6), 335-340 (1998)
   PUBMED   10048482
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp
            PROJECT = 'Dictyostelium discoideum cDNA project in Japan'
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .718
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SLC229"
                     /dev_stage="slug"
                     /clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
ORIGIN

Alignment Scores:
Pred. No.:                3.17          Length:        718
Score:                    73.50         Matches:       12
Percent Similarity:       92.31%        Conservative:  0
Best Local Similarity:    92.31%        Mismatches:    0
Query Match:              87.50%        Indels:        1
DB:                       1             Gaps:          1

US-09-858-332G-15 (1-12) x AU034181 (1-718)

Qy          1 HisAsnHisAsnHisAsn---HisAsnHisAsnHisAsn 12
              ||||||||||||||||||   ||||||||||||||||||
Db        232 CATAATCATAATCATAATCATCACAATCACAATCATAAT 270

RESULT 288
CF986668
LOCUS       CF986668                290 bp    mRNA    linear   EST 04-FEB-2005
DEFINITION  24212rsicee_5870.y1 Oryza sativa cv. LYP9 tillering whole plant
            cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF986668
VERSION     CF986668.1  GI:58588360
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 290)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 290
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .290
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="tillering"
                     /clone_lib="Oryza sativa cv. LYP9 tillering whole plant
                     cDNA library"
ORIGIN

Alignment Scores:
Pred. No.:                1.43          Length:        290
Score:                    73.00         Matches:       10
Percent Similarity:       91.67%        Conservative:  1
Best Local Similarity:    83.33%        Mismatches:    1
Query Match:              86.90%        Indels:        0
DB:                       7             Gaps:          0

US-09-858-332G-15 (1-12) x CF986668 (1-290)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||   ||||||||||||||||||||||||
Db        201 CACCACCACGAGCACAACCACAACCACAACCACAAC 236

RESULT 289
CK116691
LOCUS       CK116691                315 bp    mRNA    linear   EST 01-DEC-2003
DEFINITION  B015P72 Hybrid aspen plasmid library Populus tremula x Populus
            tremuloides cDNA clone B015P72 5', mRNA sequence.
ACCESSION   CK116691
VERSION     CK116691.1  GI:38601016
KEYWORDS    EST.
SOURCE      Populus tremula x Populus tremuloides
  ORGANISM  Populus tremula x Populus tremuloides
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Malpighiales; Salicaceae; Saliceae; Populus.
REFERENCE   1  (bases 1 to 315)
  AUTHORS   Sterky,F., Bhalerao,R.R., Unneberg,P., Segerman,B., Nilsson,P.,
            Brunner,A.M., Charbonnel-Campaa,L., Lindvall,J.J., Tandre,K.,
```

Strauss,S.H., Sundberg,B., Gustafsson,P., Uhlen,M., Bhalerao,R.P., Nilsson,O., Sandberg,G., Karlsson,J., Lundeberg,J. and Jansson,S.
TITLE A Populus EST resource for plant functional genomics
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 101 (38), 13951-13956 (2004)
PUBMED 15353603
COMMENT Contact: Bo Segerman
Umea Plant Science Center, Department of Plant Physiology
Umea University
901 87 Umea, Sweden
Tel: +46 90 786 5279
Fax: +46 90 786 6676
Email: bo.segerman@plantphys.umu.se.

FEATURES Location/Qualifiers
source 1. .315
/organism="Populus tremula x Populus tremuloides"
/mol_type="mRNA"
/db_xref="taxon:47664"
/clone="B015P72"
/tissue_type="Cambial region"
/dev_stage="1.5 m actively growing tree"
/lab_host="E.coli"
/clone_lib="Hybrid aspen plasmid library"
/note="Vector: pBluescript SK; Site_1: SalI; Site_2: NotI; Cambial region tissues, including developing xylem, the meristematic cambial zone and the developing and mature phloem, was harvested from 1.5 m actively growing trees. cDNA was prepared and cloned into lambda gt22a. DNA was isolated and subcloned into pBluescript SK using SalI and NotI restriction enzymes."

ORIGIN

Alignment Scores:

Pred. No.:	1.56	Length:	315
Score:	73.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	90.91%	Mismatches:	0
Query Match:	86.90%	Indels:	0
DB:	7	Gaps:	0

US-09-858-332G-15 (1-12) x CK116691 (1-315)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              |||:::|||||||||||||||||||||||||||
Db        167 CACCACCACAACCACAACCACAACCACAACCAC 199
```

RESULT 290
H35592
LOCUS H35592 316 bp mRNA linear EST 02-APR-1998
DEFINITION EST110865 Rat PC-12 cells, NGF-treated (9 days) Rattus norvegicus cDNA clone RPNBN84 similar to H.sapiens hypothetical protein, 13 kDa, mRNA sequence.
ACCESSION H35592
VERSION H35592.1 GI:981009
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 316)
AUTHORS Lee,N.H., Weinstock,K.G., Kirkness,E.F., Earle-Hughes,J.A., Fuldner,R.A., Marmaras,S., Glodek,A., Gocayne,J.D., Adams,M.D., Kerlavage,A.R., Fraser,C.M. and Venter,J.C.
TITLE Comparative expressed-sequence-tag analysis of differential gene expression profiles in PC-12 cells before and after nerve growth factor treatment
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 92, 8303-8307 (1995)
PUBMED 7667285
COMMENT Other_ESTs: TC725
Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
For clone availability please contact the TIGR Database (tdbinfo@tdb.tigr.org) TC (Tentative Consensus) numbers represent assemblies of ESTs.

FEATURES Location/Qualifiers
source 1. .316
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="ATCC (inhost):2005025"
/db_xref="taxon:10116"
/clone="RPNBN84"
/clone_lib="Rat PC-12 cells, NGF-treated (9 days)"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; poly(A)+ RNA was purified from 9-day NGF treated PC12 cells. cDNA was constructed using an oligo-dT primer and directionally cloned using the Lambda ZAP II Vector Kit by Stratagene"

ORIGIN

Alignment Scores:

Pred. No.:	1.56	Length:	316
Score:	73.00	Matches:	12
Percent Similarity:	85.71%	Conservative:	0
Best Local Similarity:	85.71%	Mismatches:	0
Query Match:	86.90%	Indels:	2
DB:	8	Gaps:	1

US-09-858-332G-15 (1-12) x H35592 (1-316)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHis------AsnHisAsn 12
              |||||||||||||||||||||||||||      |||||||||
Db        238 CACAACCATAACCACAACCACAACCACAGCCANAACCACAAT 279
```

RESULT 291
CF991958
LOCUS CF991958 372 bp mRNA linear EST 04-FEB-2005
DEFINITION 17088rsicee_14776.y1 Oryza sativa cv. LYP9 tillering whole plant cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA sequence.
ACCESSION CF991958
VERSION CF991958.1 GI:58593650
KEYWORDS EST.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 372)
AUTHORS Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W., Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X., Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J., Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q., Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z., Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P., Wu,S. and Liu,J.
TITLE The Genomes of Oryza sativa: A History of Duplications
JOURNAL PLoS Biol. 3 (2), e38 (2005)
PUBMED 15685292
COMMENT Contact: Yan Zhou
Bioinformatics Department
Hangzhou Genomics Institute
No.51 Zhijiang Road, Hangzhou 310008, China
Tel: 86-571-56805886
Fax: 86-571-56805884
Email: zhouyan@genomics.org.cn
Seq primer: M13 Forward
High quality sequence stop: 372
POLYA=No.

FEATURES Location/Qualifiers
source 1. .372
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"

```
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="tillering"
                     /clone_lib="Oryza sativa cv. LYP9 tillering whole plant
                     cDNA library"
ORIGIN

Alignment Scores:
Pred. No.:                1.85          Length:         372
Score:                    73.00         Matches:        10
Percent Similarity:       91.67%        Conservative:   1
Best Local Similarity:    83.33%        Mismatches:     1
Query Match:              86.90%        Indels:         0
DB:                       7             Gaps:           0

US-09-858-332G-15 (1-12) x CF991958 (1-372)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||   ||||||||||||||||||||||||
Db        193 CACCACCACGAGCACAACCACAACCACAACCACAAC 228

RESULT 292
CF966741
LOCUS       CF966741                397 bp    mRNA    linear   EST 15-JAN-2004
DEFINITION  12584rsiced_7618.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF966741
VERSION     CF966741.1  GI:38482939
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 397)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 397
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .397
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="trefoil"
                     /clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA
                     library"
ORIGIN

Alignment Scores:
Pred. No.:                1.98          Length:         397
Score:                    73.00         Matches:        10
Percent Similarity:       91.67%        Conservative:   1
Best Local Similarity:    83.33%        Mismatches:     1
Query Match:              86.90%        Indels:         0
DB:                       7             Gaps:           0

US-09-858-332G-15 (1-12) x CF966741 (1-397)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||   ||||||||||||||||||||||||
Db         63 CACCACCACGAACACAACCACAACCACAACCACAAC 98

RESULT 293
AV818700
LOCUS       AV818700                416 bp    mRNA    linear   EST 01-APR-2002
DEFINITION  AV818700 RAFL9 Arabidopsis thaliana cDNA clone RAFL09-99-F22 3',
            mRNA sequence.
ACCESSION   AV818700
VERSION     AV818700.1  GI:19860620
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 416)
  AUTHORS   Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
            Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,
            Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y.
            and Shinozaki,K.
  TITLE     Large scale analysis of Arabidopsis full-length cDNA (2002b)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            An Arabidopsis full-length cDNA library was constructed essentially
            as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
            and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et
            al., submitted for publication) digested with BamHI and SalI. This
            clone is in a modified pBluescript vector. Please visit our web
            site (http://www.gsc.riken.go.jp/e/plant/index _e.html) for further
            details.
FEATURES             Location/Qualifiers
     source          1. .416
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="RAFL09-99-F22"
                     /dev_stage="plants at various developmental stages from
                     germination to mature seeds"
                     /lab_host="DH10B"
                     /clone_lib="RAFL9"
                     /note="Site_1: BamHI; Site_2: SalI; subjected to
                     dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24
                     hr) treatments"
ORIGIN

Alignment Scores:
Pred. No.:                2.08          Length:         416
Score:                    73.00         Matches:        10
Percent Similarity:       100.00%       Conservative:   1
Best Local Similarity:    90.91%        Mismatches:     0
Query Match:              86.90%        Indels:         0
DB:                       1             Gaps:           0

US-09-858-332G-15 (1-12) x AV818700 (1-416)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              |||||||||||||||:::|||||||||||||||
Db        367 CATAATCATAATCATGATCATAACCATAATCAT 399
```

```
RESULT 294
CF096176
LOCUS       CF096176               419 bp    mRNA    linear   EST 22-JUL-2003
DEFINITION  QHN22F11.yg.ab1 QH_N sunflower H.argophyllus (drought stress)
            Helianthus argophyllus cDNA clone QHN22F11, mRNA sequence.
ACCESSION   CF096176
VERSION     CF096176.1  GI:33135243
KEYWORDS    EST.
SOURCE      Helianthus argophyllus
  ORGANISM  Helianthus argophyllus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; campanulids; Asterales; Asteraceae; Asteroideae;
            Heliantheae; Helianthus.
REFERENCE   1  (bases 1 to 419)
  AUTHORS   Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J.,
            Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y.,
            Lai,Z., Church,S., Jackson,L. and Bradford,K.
  TITLE     Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
            belongs to contig QH_CA_Contig6346, see http://cgpdb.ucdavis.edu/
            for details.
            Plate: QHN22  row: F  column: 11.
FEATURES             Location/Qualifiers
     source          1. .419
                     /organism="Helianthus argophyllus"
                     /mol_type="mRNA"
                     /db_xref="taxon:73275"
                     /clone="QHN22F11"
                     /lab_host="E.coli"
                     /clone_lib="QH_N sunflower H.argophyllus (drought stress)"
                     /note="Vector: pGEM-T; The library was constructed from
                     three different sources (seedling, root and leaf) of RNA
                     from a single genotype. cDNAs were pooled and cloned into
                     a high-copy vector pGEM-T. Details of library construction
                     can be obtained at http://cgpdb.ucdavis.edu/"
ORIGIN

Alignment Scores:
Pred. No.:                2.1          Length:          419
Score:                    73.00        Matches:         10
Percent Similarity:       100.00%      Conservative:    2
Best Local Similarity:    83.33%       Mismatches:      0
Query Match:              86.90%       Indels:          0
DB:                       6            Gaps:            0

US-09-858-332G-15 (1-12) x CF096176 (1-419)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||:::||||||:::|||
Db        289 CATAACCATAATCATAATCATCACCATAACTATAAT 324

RESULT 295
BP619865
LOCUS       BP619865               449 bp    mRNA    linear   EST 26-JUN-2004
DEFINITION  BP619865 RAFL16 Arabidopsis thaliana cDNA clone RAFL16-34-N13 3',
            mRNA sequence.
ACCESSION   BP619865
VERSION     BP619865.1  GI:49271047
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   Seki,M., Narusaka,M., Kamiya,A., Ishida,J., Satou,M., Sakurai,T.,
            Nakajima,M., Enju,A., Akiyama,K., Oono,Y., Muramatsu,M.,
            Hayashizaki,Y., Kawai,J., Carninci,P., Itoh,M., Ishii,Y.,
            Arakawa,T., Shibata,K., Shinagawa,A. and Shinozaki,K.
  TITLE     Functional annotation of a full-length Arabidopsis cDNA collection
  JOURNAL   Science 296 (5565), 141-145 (2002)
  PUBMED    11910074
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            reversed clone; Please visit our web site
            (http://pfgweb.gsc.riken.go.jp/) for further details.
FEATURES             Location/Qualifiers
     source          1. .449
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="RAFL16-34-N13"
                     /lab_host="DH10B"
                     /clone_lib="RAFL16"
                     /note="Site_1: BamHI; Site_2: SalI; dark-grown"
ORIGIN

Alignment Scores:
Pred. No.:                2.25         Length:          449
Score:                    73.00        Matches:         10
Percent Similarity:       100.00%      Conservative:    1
Best Local Similarity:    90.91%       Mismatches:      0
Query Match:              86.90%       Indels:          0
DB:                       3            Gaps:            0

US-09-858-332G-15 (1-12) x BP619865 (1-449)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 11
              ||||||||||||||||:::||||||||||||||||
Db        402 CATAATCATAATCATGATCATAACCATAATCAT 434

RESULT 296
CF985050
LOCUS       CF985050               457 bp    mRNA    linear   EST 04-FEB-2005
DEFINITION  17245rsicee_15083.y1 Oryza sativa cv. LYP9 tillering whole plant
            cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF985050
VERSION     CF985050.1  GI:58586742
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 457)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
  PUBMED    15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
```

```
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 457
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .457
                      /organism="Oryza sativa (indica cultivar-group)"
                      /mol_type="mRNA"
                      /cultivar="LYP9"
                      /db_xref="taxon:39946"
                      /tissue_type="whole plant"
                      /dev_stage="tillering"
                      /clone_lib="Oryza sativa cv. LYP9 tillering whole plant
                      cDNA library"
ORIGIN

Alignment Scores:
Pred. No.:              2.3           Length:        457
Score:                  73.00         Matches:       10
Percent Similarity:     91.67%        Conservative:  1
Best Local Similarity:  83.33%        Mismatches:    1
Query Match:            86.90%        Indels:        0
DB:                     7             Gaps:          0

US-09-858-332G-15 (1-12) x CF985050 (1-457)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||   ||||||||||||||||||||||||
Db        202 CACCACCACGAGCACAACCACAACCACAACCACAAC 237

RESULT 297
CB257174/c
LOCUS       CB257174                478 bp    mRNA    linear   EST 06-NOV-2003
DEFINITION  40-E012741-027-007-009-T7R MPIZ-ADIS-027 Arabidopsis thaliana cDNA
            clone MPIZp772O097Q 5-PRIME, mRNA sequence.
ACCESSION   CB257174
VERSION     CB257174.1  GI:32881947
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 478)
  AUTHORS   Schmid,K.J., Soerensen,T.R., Stracke,R., Torjek,O., Altmann,T.,
            Mitchell-Olds,T. and Weisshaar,B.
  TITLE     Large-scale identification and analysis of genome-wide
            single-nucleotide polymorphisms for mapping in Arabidopsis thaliana
  JOURNAL   Genome Res. 13 (6), 1250-1257 (2003)
   PUBMED   12799357
COMMENT     Contact: Weisshaar B
            ADIS DNA core facility at MPIZ
            Max-Planck-Institute for Plant Breeding Research
            Carl-von-Linne Weg 10, 50829 Koeln, Germany
            Fax: 00492215062851
            Email: weisshaa@mpiz-koeln.mpg.de
            Insert Length: 478   Std Error: 0.00
            Plate: 7  row: O  column: 09
            Seq primer: T7R; CTAATACGACTCACTATAGGGA.
FEATURES              Location/Qualifiers
     source           1. .478
                      /organism="Arabidopsis thaliana"
                      /mol_type="mRNA"
                      /ecotype="Ei-2"
                      /db_xref="GABI:592908"
                      /db_xref="taxon:3702"
                      /clone="MPIZp772O097Q"
                      /tissue_type="whole plant"
                      /dev_stage="adult plant, mixed stresses"
                      /lab_host="E. coli TOP10"
                      /clone_lib="MPIZ-ADIS-027"
                      /note="Vector: pSPORT1; Site_1: SalI; Site_2: Not1; cDNA
                      library from Arabidopsis thaliana, accession Eifel-2; ten
                      week old total plants grown under long-day conditions in
                      soil, whole adult plants were treated for 24 hours with
                      different stresses, (1) at 4M-0 C in the dark, (2), at 37
                      Grad C in the dark, (3) lying in the lab after removing
                      from soil, (4) in the greenhouse after wounding leaves
                      with a foreceps, (5) in the lab watering with a 150 mM
                      NaCl solution, (6) at 26 M-0C in the light/UV; equal
                      quantities of stressed plant material were pooled; library
                      was made at the Max-Planck-Institute for Plant Breeding
                      Research, Cologne, Germany; cloning sites SalI-NotI,
                      primer sites and orientation:
                      T7-SalI-CCACGCGTCCG-5prime-cDNA-polyA-CC-NotI-SP6; Note:
                      Sequencing granted in the context of the GABI Arabidopsis
                      Verbund I: Genetic Diversity, 'Establishment of
                      high-efficiency SNP-based mapping tools and development of
                      methods for genome-wide mutation detection' PI: Bernd
                      Weisshaar Sequence submission managed by RZPD/GABI-Primary
                      database: http://gabi.rzpd.de. This clone is available
                      from RZPD; contact RZPD (clone@rzpd.de) for further
                      information."
ORIGIN

Alignment Scores:
Pred. No.:              2.41          Length:        478
Score:                  73.00         Matches:       10
Percent Similarity:     100.00%       Conservative:  1
Best Local Similarity:  90.91%        Mismatches:    0
Query Match:            86.90%        Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-15 (1-12) x CB257174 (1-478)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              |||||||||||||||:::|||||||||||||||
Db        334 CATAATCATAATCATGATCATAACCATAATCAT 302

RESULT 298
CR888025/c
LOCUS       CR888025                525 bp    DNA     linear   GSS 22-NOV-2004
DEFINITION  Sus scrofa BES, genomic survey sequence.
ACCESSION   CR888025
VERSION     CR888025.1  GI:55966689
KEYWORDS    GSS; Bac-end sequence BES; Genome Survey Sequence.
SOURCE      Sus scrofa (pig)
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
            Sus.
REFERENCE   1  (bases 1 to 525)
  AUTHORS   Rogel-Gaillard,C., Bourgeaux,N., Billault,A., Vaiman,M. and
            Chardon,P.
  TITLE     Construction of a swine BAC library: application to the
            characterization and mapping of porcine type C endoviral elements
  JOURNAL   Cytogenet. Cell Genet. 85 (3-4), 205-211 (1999)
   PUBMED   10449899
REFERENCE   2  (bases 1 to 525)
  AUTHORS   Chardon,P., Iannuccelli,N., Roig,A., Dossat,C., Demars,J.,
            Rogel-Gaillard,C., Roy,A., Schibler,L. and Milan,D.
  TITLE     A physical map of the swine genome
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 525)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
FEATURES              Location/Qualifiers
     source           1. .525
                      /organism="Sus scrofa"
                      /mol_type="genomic DNA"
                      /strain="Large White"
                      /db_xref="taxon:9823"
```

/clone="bI0129A07"
/sex="male"
/cell_type="fibroblast"
/clone_lib="SBAB"
/note="Genoscope sequence ID : IH0AAA11BB03FM1"

ORIGIN

Alignment Scores:

Pred. No.:	2.65	Length:	525
Score:	73.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	90.91%	Mismatches:	0
Query Match:	86.90%	Indels:	0
DB:	11	Gaps:	0

US-09-858-332G-15 (1-12) x CR888025 (1-525)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              |||:::|||||||||||||||||||||||||||
Db        227 CACAGCCACAACCACAACCACAACCACAACCAC 195
```

RESULT 299
CB257064/c
LOCUS CB257064 591 bp mRNA linear EST 06-NOV-2003
DEFINITION 47-E012741-027-007-M11-T7R MPIZ-ADIS-027 Arabidopsis thaliana cDNA clone MPIZp772M117Q 5-PRIME, mRNA sequence.
ACCESSION CB257064
VERSION CB257064.1 GI:32881837
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 591)
AUTHORS Schmid,K.J., Soerensen,T.R., Stracke,R., Torjek,O., Altmann,T., Mitchell-Olds,T. and Weisshaar,B.
TITLE Large-scale identification and analysis of genome-wide single-nucleotide polymorphisms for mapping in Arabidopsis thaliana
JOURNAL Genome Res. 13 (6), 1250-1257 (2003)
PUBMED 12799357
COMMENT Contact: Weisshaar B
ADIS DNA core facility at MPIZ
Max-Planck-Institute for Plant Breeding Research
Carl-von-Linne Weg 10, 50829 Koeln, Germany
Fax: 00492215062851
Email: weisshaa@mpiz-koeln.mpg.de
Insert Length: 591 Std Error: 0.00
Plate: 7 row: M column: 11
Seq primer: T7R; CTAATACGACTCACTATAGGGA.
FEATURES Location/Qualifiers
source 1. .591
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/ecotype="Ei-2"
/db_xref="GABI:593032"
/db_xref="taxon:3702"
/clone="MPIZp772M117Q"
/tissue_type="whole plant"
/dev_stage="adult plant, mixed stresses"
/lab_host="E. coli TOP10"
/clone_lib="MPIZ-ADIS-027"
/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; cDNA library from Arabidopsis thaliana, accession Eifel-2; ten week old total plants grown under long-day conditions in soil, whole adult plants were treated for 24 hours with different stresses, (1) at 4M-0 C in the dark, (2), at 37 Grad C in the dark, (3) lying in the lab after removing from soil, (4) in the greenhouse after wounding leaves with a foreceps, (5) in the lab watering with a 150 mM NaCl solution, (6) at 26 M-0C in the light/UV; equal quantities of stressed plant material were pooled; library was made at the Max-Planck-Institute for Plant Breeding Research, Cologne, Germany; cloning sites SalI-NotI, primer sites and orientation: T7-SalI-CCACGCGTCCG-5prime-cDNA-polyA-CC-NotI-SP6; Note: Sequencing granted in the context of the GABI Arabidopsis Verbund I: Genetic Diversity, 'Establishment of high-efficiency SNP-based mapping tools and development of methods for genome-wide mutation detection' PI: Bernd Weisshaar Sequence submission managed by RZPD/GABI-Primary database: http://gabi.rzpd.de. This clone is available from RZPD; contact RZPD (clone@rzpd.de) for further information."

ORIGIN

Alignment Scores:

Pred. No.:	3	Length:	591
Score:	73.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	90.91%	Mismatches:	0
Query Match:	86.90%	Indels:	0
DB:	6	Gaps:	0

US-09-858-332G-15 (1-12) x CB257064 (1-591)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              ||||||||||||||:::|||||||||||||||||
Db        445 CATAATCATAATCATGATCATAACCATAATCAT 413
```

RESULT 300
AJ803321
LOCUS AJ803321 667 bp mRNA linear EST 08-DEC-2004
DEFINITION AJ803321 Antirrhinum majus whole plant Antirrhinum majus cDNA clone 018_5_06_j01, mRNA sequence.
ACCESSION AJ803321
VERSION AJ803321.1 GI:51118649
KEYWORDS EST.
SOURCE Antirrhinum majus (snapdragon)
ORGANISM Antirrhinum majus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; lamiids; Lamiales; Plantaginaceae; Antirrhineae; Antirrhinum.
REFERENCE 1 (bases 1 to 667)
AUTHORS Bey,M., Stueber,K., Fellenberg,K., Schwarz-Sommer,Z., Sommer,H., Saedler,H. and Zachgo,S.
TITLE Characterization of Antirrhinum Petal Development and Identification of Target Genes of the Class B MADS Box Gene DEFICIENS
JOURNAL Plant Cell 16 (12), 3197-3215 (2004)
PUBMED 15539471
COMMENT Contact: Schwarz-Sommer Z
Molekulare Pflanzengenetik
MPI fuer Zuechtungsforschung
Carl-von-Linne Weg 10, D-50829, Germany.
FEATURES Location/Qualifiers
source 1. .667
/organism="Antirrhinum majus"
/mol_type="mRNA"
/db_xref="taxon:4151"
/clone="018_5_06_j01"
/tissue_type="whole plant"
/clone_lib="Antirrhinum majus whole plant"

ORIGIN

Alignment Scores:

Pred. No.:	3.41	Length:	667
Score:	73.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	90.91%	Mismatches:	0
Query Match:	86.90%	Indels:	0
DB:	1	Gaps:	0

US-09-858-332G-15 (1-12) x AJ803321 (1-667)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              ||||||||||||||||||||:::|||||||||
Db         69 CACAACCACAATCACAACCACAGCCACAACCAC 101
```

Search completed: November 28, 2005, 11:00:22
Job time : 2652 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 28, 2005, 08:20:11 ; Search time 3044 Seconds
(without alignments)
224.087 Million cell updates/sec

Title: US-09-858-332G-15
Perfect score: 84
Sequence: 1 HNHNHNHNHNHN 12

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 600 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09858332/runat_23112005_103443_19713/app_query.fasta_1.199
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=600
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=300 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09858332_@CGN_1_1_4939_@runat_23112005_103443_19713 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	84	100.0	42	6	AX594160	AX594160 Sequence
	2	84	100.0	100	6	AX594158	AX594158 Sequence
c	3	84	100.0	115	4	AY559840	AY559840 Bos tauru
c	4	84	100.0	115	6	AR568829	AR568829 Sequence
c	5	84	100.0	387	4	AF097207	AF097207 Bos tauru
c	6	84	100.0	390	4	AF097209	AF097209 Bos tauru
	7	84	100.0	427	2	PFA430088	AJ430088 Plasmodiu
	8	84	100.0	431	10	BV211484	BV211484 BARC1076
c	9	84	100.0	478	4	BTU49779	U49779 Bos taurus
	10	84	100.0	519	9	AF187065	AF187065 Rattus no
	11	84	100.0	887	9	BC058503	BC058503 Rattus no
c	12	84	100.0	994	10	CNS06HW9	AL399535 T3 end of
c	13	84	100.0	1132	15	AY065142	AY065142 Arabidops
	14	84	100.0	1704	15	AK107456	AK107456 Oryza sat
	15	84	100.0	1897	6	I34914	I34914 Sequence 1
	16	84	100.0	1897	15	YSCOPI1	M57383 Saccharomyc
	17	84	100.0	1898	6	AX536592	AX536592 Sequence
	18	84	100.0	1949	9	BC029741	BC029741 Mus muscu
	19	84	100.0	2011	9	AB039933	AB039933 Mus muscu
	20	84	100.0	2152	9	BC083830	BC083830 Rattus no
c	21	84	100.0	2270	6	AX461526	AX461526 Sequence
	22	84	100.0	2396	2	AF020407	AF020407 Dictyoste
	23	84	100.0	4938	6	AX594148	AX594148 Sequence
	24	84	100.0	6525	6	AX594150	AX594150 Sequence
	25	84	100.0	7487	6	AX594151	AX594151 Sequence
	26	84	100.0	7801	2	AY160101	AY160101 Dictyoste
	27	84	100.0	9024	2	AF411930	AF411930 Plasmodiu
	28	84	100.0	9044	2	AF411933	AF411933 Plasmodiu
	29	84	100.0	9049	2	AF411929	AF411929 Plasmodiu
	30	84	100.0	20247	9	AY520825	AY520825 Mus muscu
c	31	84	100.0	29276	15	YSCH9433	U11582 Saccharomyc
	32	84	100.0	32083	15	AF309805	AF309805 Pneumocys
c	33	84	100.0	64707	2	AC115607	AC115607 Dictyoste
	34	84	100.0	65263	14	AC113185	AC113185 Mus muscu
	35	84	100.0	75741	15	AC149204	AC149204 Medicago
c	36	84	100.0	79776	14	AC139431	AC139431 Homo sapi
	37	84	100.0	83029	15	ATT8O5	AL021890 Arabidops
c	38	84	100.0	101256	15	CNS07TIW	AL627322 DNA centr
	39	84	100.0	101752	15	AP006378	AP006378 Lotus cor
c	40	84	100.0	110000	14	AC109729_3	Continuation (4 of
c	41	84	100.0	110000	14	AC120234_1	Continuation (2 of
c	42	84	100.0	110000	14	AC141230_3	Continuation (4 of
c	43	84	100.0	110000	15	AP008214_000	AP008214 Oryza sat
c	44	84	100.0	110000	15	CR382132_16	Continuation (17 o
c	45	84	100.0	110000	15	AE016820_05	Continuation (6 of
c	46	84	100.0	110000	15	AE017344_16	Continuation (17 o
	47	84	100.0	110560	14	AP007372	AP007372 Lotus cor
c	48	84	100.0	112057	14	AP004065	AP004065 Oryza sat
c	49	84	100.0	122984	14	AC151482	AC151482 Medicago
c	50	84	100.0	125310	15	AC157473	AC157473 Medicago
	51	84	100.0	125623	2	AC115599	AC115599 Dictyoste
c	52	84	100.0	125958	2	AC115592	AC115592 Dictyoste
c	53	84	100.0	127902	15	AP005406	AP005406 Oryza sat
c	54	84	100.0	144092	5	BX465845	BX465845 Zebrafish
c	55	84	100.0	153751	2	AC116551	AC116551 Dictyoste
c	56	84	100.0	155983	8	AY884207	AY884207 Homo sapi
c	57	84	100.0	157087	14	AC142155	AC142155 Rattus no
	58	84	100.0	157193	8	AL355315	AL355315 Human DNA
	59	84	100.0	161701	14	AC117110	AC117110 Rattus no
c	60	84	100.0	161965	9	AC148008	AC148008 Mus muscu
c	61	84	100.0	163087	14	AC139383	AC139383 Rattus no
	62	84	100.0	168456	9	AC159252	AC159252 Mus muscu
c	63	84	100.0	170533	8	AP003356	AP003356 Homo sapi
c	64	84	100.0	176733	14	AC068001	AC068001 Homo sapi
c	65	84	100.0	179415	9	AC139578	AC139578 Mus muscu
c	66	84	100.0	179452	14	AC164547	AC164547 Mus muscu
	67	84	100.0	182870	2	AC116960	AC116960 Dictyoste
c	68	84	100.0	182871	2	AC117176	AC117176 Dictyoste
c	69	84	100.0	184541	8	AC011405	AC011405 Homo sapi
	70	84	100.0	184741	9	AC116475	AC116475 Mus muscu
c	71	84	100.0	188453	9	AC123741	AC123741 Mus muscu
c	72	84	100.0	189130	9	AC132110	AC132110 Mus muscu
	73	84	100.0	189505	9	AC134540	AC134540 Mus muscu
c	74	84	100.0	189996	9	AC131664	AC131664 Mus muscu
c	75	84	100.0	190345	14	CR450684	CR450684 Danio rer
c	76	84	100.0	191308	8	AC034243	AC034243 Homo sapi

c	77	84	100.0	191824	9	AC159207
	78	84	100.0	194916	15	ATCHRIV55
	79	84	100.0	196247	15	ATCHRIV56
c	80	84	100.0	198321	9	AC104863
	81	84	100.0	198341	9	AC122454
	82	84	100.0	201932	5	BX897729
c	83	84	100.0	204940	14	AC069007
c	84	84	100.0	207611	8	AC091132
c	85	84	100.0	210631	9	AC120797
	86	84	100.0	218133	8	AC074051
	87	84	100.0	220006	14	AC084068
c	88	84	100.0	226013	14	AC073680
c	89	84	100.0	227081	9	AC119847
	90	84	100.0	227648	14	AC133743
	91	84	100.0	228353	9	AC154121
	92	84	100.0	228485	14	CT009527
c	93	84	100.0	231383	14	AC096973
	94	84	100.0	231384	14	AC159263
	95	84	100.0	232413	14	AC117096
	96	84	100.0	235546	14	CR450845
c	97	84	100.0	237732	14	AC127135
	98	84	100.0	240663	14	AC097541
c	99	84	100.0	240773	14	AC106512
c	100	84	100.0	242626	14	AC152062
	101	84	100.0	243772	14	AY491413
c	102	84	100.0	248752	14	AC094502
c	103	84	100.0	248784	14	AC113885
c	104	84	100.0	250591	14	AC097616
	105	84	100.0	252619	14	AC095012
	106	84	100.0	253494	14	AC126742
	107	84	100.0	254733	2	AC117075
	108	84	100.0	255032	14	AC107178
	109	84	100.0	255050	14	AC112305
c	110	84	100.0	256033	14	AC151980
c	111	84	100.0	256717	14	AC114024
	112	84	100.0	258253	14	AC136422
	113	84	100.0	260849	14	AC119447
	114	84	100.0	262405	14	AC102976
	115	84	100.0	264615	14	AC121457
c	116	84	100.0	268063	14	AC110912
	117	84	100.0	268644	14	AC091323
	118	84	100.0	269740	14	AC133056
	119	84	100.0	273729	14	AC106147
c	120	84	100.0	274368	14	AC115186
	121	84	100.0	286094	14	AC126663
	122	84	100.0	287586	14	AC103294
	123	84	100.0	295223	14	AC120291
	124	84	100.0	300143	1	AE017165
	125	84	100.0	322194	15	CNS09S4S
	126	79	94.0	72409	14	AC131257
c	127	79	94.0	110000	15	CR380953_03
c	128	79	94.0	110000	15	CR382127_10
c	129	79	94.0	110000	15	CR382127_11
c	130	79	94.0	135637	8	AP003400
	131	79	94.0	164631	14	AC163335
c	132	79	94.0	168425	14	AC012198
c	133	79	94.0	177345	9	AC132952
c	134	79	94.0	188256	14	AP001927
	135	79	94.0	200563	9	AC151575
	136	79	94.0	202116	14	AC166819
	137	79	94.0	222745	14	AC127100
c	138	79	94.0	260910	14	AC112891
	139	79	94.0	323443	14	AC145086
c	140	78	92.9	139	6	AR568831
c	141	78	92.9	205	4	AY559839
c	142	78	92.9	205	6	AR568828
c	143	78	92.9	211	4	BTA512977
c	144	78	92.9	244	15	AY095308
c	145	78	92.9	263	15	AY095306
c	146	78	92.9	267	15	AY095307
c	147	78	92.9	381	4	AF097206
c	148	78	92.9	408	4	BTU49781
c	149	78	92.9	480	4	BTU50559

AC159207 Mus muscu
AL161555 Arabidops
AL161556 Arabidops
AC104863 Mus muscu
AC122454 Mus muscu
BX897729 Zebrafish
AC069007 Homo sapi
AC091132 Homo sapi
AC120797 Mus muscu
AC074051 Homo sapi
AC084068 Mus muscu
AC073680 Mus muscu
AC119847 Mus muscu
AC133743 Rattus no
AC154121 Mus muscu
CT009527 Mus muscu
AC096973 Rattus no
AC159263 Mus muscu
AC117096 Rattus no
CR450845 Danio rer
AC127135 Rattus no
AC097541 Rattus no
AC106512 Rattus no
AC152062 Mus muscu
AY491413 Mus muscu
AC094502 Rattus no
AC113885 Rattus no
AC097616 Rattus no
AC095012 Rattus no
AC126742 Rattus no
AC117075 Dictyoste
AC107178 Rattus no
AC112305 Rattus no
AC151980 Mus muscu
AC114024 Rattus no
AC136422 Rattus no
AC119447 Rattus no
AC102976 Rattus no
AC121457 Rattus no
AC110912 Mus muscu
AC091323 Mus muscu
AC133056 Rattus no
AC106147 Rattus no
AC115186 Rattus no
AC126663 Rattus no
AC103294 Rattus no
AC120291 Rattus no
AE017165 Prochloro
BX088700 DNA centr
AC131257 Homo sapi
Continuation (4 of
Continuation (11 o
Continuation (12 o
AP003400 Homo sapi
AC163335 Mus muscu
AC012198 Homo sapi
AC132952 Mus muscu
AP001927 Homo sapi
AC151575 Mus muscu
AC166819 Mus muscu
AC127100 Rattus no
AC112891 Rattus no
AC145086 Mus muscu
AR568831 Sequence
AY559839 Bos tauru
AR568828 Sequence
AJ512977 Bos tauru
AY095308 Amaranthu
AY095306 Amaranthu
AY095307 Amaranthu
AF097206 Bos tauru
U49781 Bos taurus
U50559 Bos taurus

c	150	78	92.9	489	4	AF097203
c	151	78	92.9	492	4	AF097204
c	152	78	92.9	587	6	AR568827
	153	78	92.9	726	6	AR546135
c	154	78	92.9	1731	6	AR449702
	155	78	92.9	2580	6	AR527958
	156	78	92.9	3455	6	AR146444
	157	78	92.9	3455	6	AR579950
	158	78	92.9	3455	15	YSCMPT5
	159	78	92.9	3769	15	YSCHTR1
	160	78	92.9	5378	15	SCYGL176C
	161	78	92.9	6323	15	SCYGL178W
	162	78	92.9	7951	15	SCDNAGMH
c	163	78	92.9	109787	15	AC126793
	164	78	92.9	110000	15	CR382128_04
c	165	78	92.9	110000	15	AP008209_138
c	166	78	92.9	110000	15	AP008209_139
	167	78	92.9	122576	15	NCB13H18
	168	78	92.9	161676	15	AC135907
	169	78	92.9	191497	9	AC116420
	170	78	92.9	215707	14	AC099096
c	171	78	92.9	236844	14	AC098096
	172	78	92.9	249236	14	AC118186
	173	78	92.9	249878	6	AX770903
	174	78	92.9	254123	14	AC125668
c	175	78	92.9	349071	1	BX571864
	176	77	91.7	921	6	AR551978
	177	77	91.7	3379	2	AF111943
c	178	77	91.7	93305	14	AC138752
c	179	77	91.7	107217	8	AC003070
	180	77	91.7	110000	14	AC118389_4
	181	77	91.7	110000	14	CT009752_5
c	182	77	91.7	134798	5	AC146502
	183	77	91.7	142462	8	AC079238
c	184	77	91.7	146570	2	AC117076
	185	77	91.7	160634	14	AC141999
c	186	77	91.7	181948	5	AC146504
c	187	77	91.7	191642	5	CR524481
	188	77	91.7	202314	14	AC132694
	189	77	91.7	218242	14	AC123276
c	190	77	91.7	226154	14	AC136813
	191	77	91.7	229478	14	AC151871
	192	77	91.7	233887	14	AC133687
	193	77	91.7	247726	14	AC128004
	194	77	91.7	251231	14	AC095505
	195	77	91.7	253932	14	AC120488
	196	77	91.7	265001	14	AC121746
c	197	77	91.7	287652	14	AC118765
	198	77	91.7	331039	2	AC116979
	199	77	91.7	333321	2	AC116986
	200	76	90.5	122	4	AJ811567
c	201	76	90.5	390	4	AF097205
	202	76	90.5	537	6	A57711
	203	76	90.5	912	6	A57713
	204	76	90.5	1154	6	AX164109
	205	76	90.5	1172	6	AX164111
	206	76	90.5	1521	6	AR578923
	207	76	90.5	1521	6	AX240769
	208	76	90.5	1521	6	AX488745
	209	76	90.5	1806	6	AR545053
	210	76	90.5	3897	15	AF332954
	211	76	90.5	4167	15	AF332955
c	212	76	90.5	4565	6	DD013389
c	213	76	90.5	4565	6	AR542626
c	214	76	90.5	4565	6	AX235897
	215	76	90.5	4565	15	ZMAC1
c	216	76	90.5	4565	15	ZMTAC
c	217	76	90.5	4810	15	MZEAC9WX
	218	76	90.5	5772	15	AF355378
	219	76	90.5	6875	11	AF433042
	220	76	90.5	6928	11	AF433043
	221	76	90.5	7635	11	AB055064
	222	76	90.5	8724	4	CDR131945

AF097203 Bos tauru
AF097204 Bos tauru
AR568827 Sequence
AR546135 Sequence
AR449702 Sequence
AR527958 Sequence
AR146444 Sequence
AR579950 Sequence
D26184 Yeast MPT5
D25541 Saccharomyc
Z72698 S.cerevisia
Z72700 S.cerevisia
X83690 S.cerevisia
AC126793 Medicago
Continuation (5 of
Continuation (139
Continuation (140
AL807374 Neurospor
AC135907 Genomic s
AC116420 Mus muscu
AC099096 Rattus no
AC098096 Rattus no
AC118186 Rattus no
AX770903 Sequence
AC125668 Rattus no
BX571864 Photorhab
AR551978 Sequence
AF111943 Dictyoste
AC138752 Homo sapi
AC003070 Homo sapi
Continuation (5 of
Continuation (6 of
AC146502 Danio rer
AC079238 Homo sapi
AC117076 Dictyoste
AC141999 Rattus no
AC146504 Danio rer
CR524481 Zebrafish
AC132694 Rattus no
AC123276 Rattus no
AC136813 Rattus no
AC151871 Lemur cat
AC133687 Rattus no
AC128004 Rattus no
AC095505 Rattus no
AC120488 Rattus no
AC121746 Rattus no
AC118765 Rattus no
AC116979 Dictyoste
AC116986 Dictyoste
AJ811567 Lama glam
AF097205 Bos tauru
A57711 Sequence 21
A57713 Sequence 23
AX164109 Sequence
AX164111 Sequence
AR578923 Sequence
AX240769 Sequence
AX488745 Sequence
AR545053 Sequence
AF332954 Zea mays
AF332955 Zea mays
DD013389 Methods a
AR542626 Sequence
AX235897 Sequence
X05424 Maize trans
X01380 Zea mays tr
K01964 Maize trans
AF355378 Zea mays
AF433042 Cloning v
AF433043 Cloning v
AB055064 Synthetic
AJ131945 Camelus d

	223	76	90.5	8867	15	AY191010 AY191010 Zea mays
c	224	76	90.5	28708	11	AY218787 AY218787 Cloning v
c	225	76	90.5	110000	2	AC116305_1 Continuation (2 of
c	226	76	90.5	111067	5	BX663527 BX663527 Chicken D
	227	76	90.5	169794	14	AC004688 AC004688 Plasmodiu
	228	76	90.5	173264	9	AC151411 AC151411 Mus muscu
c	229	76	90.5	177414	14	AC097154 AC097154 Rattus no
	230	76	90.5	178733	14	AC068786 AC068786 Homo sapi
c	231	76	90.5	186190	9	AC121770 AC121770 Mus muscu
c	232	76	90.5	191770	9	AC147567 AC147567 Mus muscu
	233	76	90.5	196149	14	AC004709 AC004709 Plasmodiu
	234	76	90.5	197819	9	AC140676 AC140676 Mus muscu
c	235	76	90.5	215675	14	AC123264 AC123264 Rattus no
	236	76	90.5	220526	9	AC157594 AC157594 Mus muscu
c	237	76	90.5	232340	14	AC165427 AC165427 Mus muscu
	238	76	90.5	241828	14	AC095709 AC095709 Rattus no
	239	76	90.5	252650	2	AE014847 AE014847 Plasmodiu
c	240	76	90.5	336001	14	AC098369 AC098369 Rattus no
c	241	75	89.3	148600	8	AC002367 AC002367 Homo sapi
c	242	75	89.3	165738	14	AC019165 AC019165 Homo sapi
	243	75	89.3	194447	8	AC150208 AC150208 Pan trogl
	244	75	89.3	207665	8	AC150209 AC150209 Pan trogl
	245	74	88.1	465	2	DVNEURS1 U12591 Drosophila
	246	74	88.1	467	10	BV004681 BV004681 S208P6706
	247	74	88.1	747	6	AR546771 AR546771 Sequence
	248	74	88.1	5687	2	AY122262 AY122262 Drosophil
c	249	74	88.1	8167	2	AF474341 AF474341 Dictyoste
	250	74	88.1	42273	15	NCB12N19 AL669987 Neurospor
c	251	74	88.1	45870	2	AC024772 AC024772 Caenorhab
	252	74	88.1	110000	15	CR382127_05 Continuation (6 of
	253	74	88.1	148802	2	AC159703 AC159703 Trypanoso
c	254	74	88.1	156552	5	CR626888 CR626888 Zebrafish
c	255	74	88.1	160655	8	AC064874 AC064874 Homo sapi
c	256	74	88.1	164873	2	AC159405 AC159405 Trypanoso
	257	74	88.1	181206	9	AC119200 AC119200 Mus muscu
	258	74	88.1	184783	9	AC133201 AC133201 Mus muscu
	259	74	88.1	198631	9	AL683804 AL683804 Mouse DNA
c	260	74	88.1	199917	14	AC006755 AC006755 Caenorhab
	261	74	88.1	209398	14	AC101656 AC101656 Mus muscu
c	262	74	88.1	222379	9	AC148987 AC148987 Mus muscu
c	263	74	88.1	222776	14	AC103448 AC103448 Rattus no
	264	74	88.1	228998	14	AC112472 AC112472 Rattus no
c	265	74	88.1	238494	14	AC095717 AC095717 Rattus no
	266	73.5	87.5	2400	2	AF036170 AF036170 Dictyoste
	267	73.5	87.5	5284	2	DDDOKAGEN X96869 D.discoideu
c	268	73.5	87.5	110000	2	AC116957_0 AC116957 Dictyoste
c	269	73.5	87.5	254816	14	AC094631 AC094631 Rattus no
c	270	73	86.9	862	15	AY084699 AY084699 Arabidops
	271	73	86.9	14923	14	AC013163 AC013163 Drosophil
c	272	73	86.9	39074	14	AC156596 AC156596 Laccaria
c	273	73	86.9	71813	2	DMBN4L24 AL121805 Drosophil
c	274	73	86.9	85992	15	AB009052 AB009052 Arabidops
c	275	73	86.9	88549	2	AC116924 AC116924 Dictyoste
c	276	73	86.9	99967	15	NCB23B10 BX284752 Neurospor
c	277	73	86.9	110000	2	AE003426_2 Continuation (3 of
	278	73	86.9	179069	14	AC104144 AC104144 Drosophil
	279	73	86.9	182387	2	AC116543 AC116543 Drosophil
	280	73	86.9	221040	14	AC097439 AC097439 Rattus no
c	281	73	86.9	221938	14	AC098906 AC098906 Rattus no
	282	73	86.9	248646	14	AC113650 AC113650 Rattus no
c	283	73	86.9	254668	14	AC132543 AC132543 Rattus no
c	284	73	86.9	254733	2	AC117075 AC117075 Dictyoste
c	285	72	85.7	190	4	BTA512824 AJ512824 Bos tauru
c	286	72	85.7	209	4	BTA512976 AJ512976 Bos tauru
c	287	72	85.7	360	4	BTU49777 U49777 Bos taurus
	288	72	85.7	1239	5	DQ120961 DQ120961 Hucho tai
	289	72	85.7	1953	6	AR549509 AR549509 Sequence
	290	72	85.7	3042	6	AX536944 AX536944 Sequence
c	291	72	85.7	5826	6	AX345465 AX345465 Sequence
c	292	72	85.7	6236	6	AX344794 AX344794 Sequence
c	293	72	85.7	6236	6	AX346532 AX346532 Sequence
c	294	72	85.7	39156	15	AC165067 AC165067 Phakopsor
c	295	72	85.7	40374	5	AL845551 AL845551 Zebrafish
c	296	72	85.7	81767	14	AC021929 AC021929 Homo sapi
c	297	72	85.7	110000	15	CR380953_01 Continuation (2 of
c	298	72	85.7	110000	15	CR382129_00 CR382129 Yarrowia
c	299	72	85.7	122102	8	AC010468 AC010468 Homo sapi
	300	72	85.7	161081	5	CR855273 CR855273 Zebrafish
	301	72	85.7	161452	8	AC020925 AC020925 Homo sapi
c	302	72	85.7	166604	9	AC125168 AC125168 Mus muscu
c	303	72	85.7	182733	14	AC103609 AC103609 Mus muscu
	304	72	85.7	188630	14	AC012606 AC012606 Homo sapi
	305	72	85.7	239036	14	AC103112 AC103112 Rattus no
c	306	72	85.7	256879	2	AC116982 AC116982 Dictyoste
c	307	72	85.7	282366	14	AC103237 AC103237 Rattus no
c	308	72	85.7	285677	14	AC131858 AC131858 Rattus no
	309	72	85.7	332460	14	AC137373 AC137373 Rattus no
c	310	72	85.7	349980	6	AX344563 AX344563 Sequence
	311	71.5	85.1	2790	2	DDI301670 AJ301670 Dictyoste
	312	71	84.5	536	15	AY677216 AY677216 Campyloca
	313	71	84.5	536	15	AY677217 AY677217 Campyloca
	314	71	84.5	536	15	AY677218 AY677218 Campyloca
	315	71	84.5	536	15	AY677219 AY677219 Campyloca
	316	71	84.5	536	15	AY677220 AY677220 Campyloca
	317	71	84.5	536	15	AY677221 AY677221 Campyloca
	318	71	84.5	553	15	AY677222 AY677222 Campyloca
	319	71	84.5	553	15	AY677223 AY677223 Campyloca
	320	71	84.5	553	15	AY677224 AY677224 Campyloca
	321	71	84.5	557	15	AY677225 AY677225 Campyloca
	322	71	84.5	903	6	AR549211 AR549211 Sequence
c	323	71	84.5	17284	9	AL672209 AL672209 Mouse DNA
	324	71	84.5	74037	8	HS1071N3 AL031728 Human DNA
	325	71	84.5	112516	8	AC080079 AC080079 Homo sapi
c	326	71	84.5	131860	6	AR659689 AR659689 Sequence
	327	71	84.5	136188	14	AC160111 AC160111 Mus muscu
	328	71	84.5	158807	14	AC011771 AC011771 Homo sapi
	329	71	84.5	163443	14	AC006280 AC006280 Plasmodiu
	330	71	84.5	176787	14	AC021374 AC021374 Homo sapi
	331	71	84.5	177298	14	AC068751 AC068751 Homo sapi
	332	71	84.5	182896	14	AC011043 AC011043 Homo sapi
c	333	71	84.5	190048	9	AL591126 AL591126 Mouse DNA
	334	71	84.5	205429	14	AC005506 AC005506 Plasmodiu
c	335	71	84.5	214284	14	AC112544 AC112544 Rattus no
	336	71	84.5	219712	14	AC127196 AC127196 Rattus no
	337	71	84.5	221265	14	CT025587 CT025587 Mus muscu
	338	71	84.5	231647	14	AC114203 AC114203 Rattus no
	339	71	84.5	244147	14	AC094218 AC094218 Rattus no
	340	71	84.5	252554	14	AC115127 AC115127 Rattus no
c	341	71	84.5	253132	2	AE014846 AE014846 Plasmodiu
	342	70	83.3	192	6	AR553849 AR553849 Sequence
	343	70	83.3	601	6	AR664528 AR664528 Sequence
	344	70	83.3	601	6	AR664529 AR664529 Sequence
	345	70	83.3	930	15	CAL390515 AJ390515 Candida a
	346	70	83.3	966	6	AR545979 AR545979 Sequence
	347	70	83.3	1339	15	AJ621916 AJ621916 Pisum sat
	348	70	83.3	1709	2	S55235 S55235 cAR3=cAMP r
	349	70	83.3	1995	6	AR557473 AR557473 Sequence
c	350	70	83.3	2399	2	DMU66562 U66562 Drosophila
c	351	70	83.3	2481	2	DMU67304 U67304 Drosophila
	352	70	83.3	2527	15	AY502069 AY502069 Pisum sat
	353	70	83.3	2545	15	AY502068 AY502068 Pisum sat
c	354	70	83.3	2556	6	A72166 A72166 Sequence 1
c	355	70	83.3	2556	6	AR616818 AR616818 Sequence
	356	70	83.3	2574	15	AY502067 AY502067 Pisum sat
	357	70	83.3	2586	6	CS128727 CS128727 Sequence
	358	70	83.3	2586	6	CS135509 CS135509 Sequence
	359	70	83.3	5499	2	AF068065 AF068065 Cryptospo
	360	70	83.3	5511	6	AR097040 AR097040 Sequence
	361	70	83.3	5511	6	AR279259 AR279259 Sequence
	362	70	83.3	7334	6	AR097039 AR097039 Sequence
	363	70	83.3	7334	6	AR279258 AR279258 Sequence
	364	70	83.3	24076	6	AX537040 AX537040 Sequence
	365	70	83.3	56494	2	AE003846_3 Continuation (4 of
c	366	70	83.3	60713	14	AC073922 AC073922 Homo sapi
c	367	70	83.3	64958	14	AC090905 AC090905 Homo sapi
	368	70	83.3	74509	14	AF260012 AF260012 Homo sapi

	369	70	83.3	75148	8	AC091545
	370	70	83.3	84539	14	AC106241_5
	371	70	83.3	99914	5	BX890595
	372	70	83.3	100000	8	AP000093
	373	70	83.3	100000	8	AP000197
	374	70	83.3	103979	5	CR352288
	375	70	83.3	110000	2	AC116305_3
	376	70	83.3	110000	14	AC106241_4
c	377	70	83.3	110000	14	AC110363_4
	378	70	83.3	110000	14	AC111637_2
	379	70	83.3	110000	15	CR380947_1
	380	70	83.3	110000	15	AP008209_255
c	381	70	83.3	116874	5	CR383661
c	382	70	83.3	117046	15	AC135600
	383	70	83.3	146478	14	AC032036
c	384	70	83.3	149253	8	AP002001
c	385	70	83.3	152828	15	AC133859
c	386	70	83.3	153265	8	AP000237
c	387	70	83.3	157226	8	AC018902
	388	70	83.3	157672	8	AC113143
	389	70	83.3	160496	14	AC164419
	390	70	83.3	160746	14	AC044853
	391	70	83.3	161940	14	AC148232
c	392	70	83.3	161972	8	AP001985
	393	70	83.3	167578	5	BX649370
c	394	70	83.3	172780	9	AC117186
c	395	70	83.3	176748	9	AC102446
c	396	70	83.3	181459	14	AC053477
c	397	70	83.3	183358	8	AC019041
c	398	70	83.3	183590	9	AC123713
c	399	70	83.3	194229	9	AC140214
	400	70	83.3	195419	5	BX000486
	401	70	83.3	198087	8	AC008074
c	402	70	83.3	199539	9	AC121818
	403	70	83.3	199827	9	AC137110
c	404	70	83.3	208412	9	AC090977
	405	70	83.3	210968	9	AC113022
	406	70	83.3	211693	14	AC105714
	407	70	83.3	212251	9	AC123977
	408	70	83.3	216702	14	AC130099
c	409	70	83.3	220614	14	AC094779
	410	70	83.3	221974	14	AC099118
c	411	70	83.3	229606	14	AC105519
c	412	70	83.3	231284	14	AC109415
c	413	70	83.3	231593	9	AC140433
c	414	70	83.3	232792	14	AC095568
	415	70	83.3	233077	9	AL807771
	416	70	83.3	237533	14	AC113874
c	417	70	83.3	241925	9	AC102528
	418	70	83.3	242803	14	AC108557
c	419	70	83.3	247144	14	AC111654
c	420	70	83.3	253430	14	AC095305
	421	70	83.3	258571	14	AC097144
	422	70	83.3	263507	14	AC093466
c	423	70	83.3	320775	14	AC126201
c	424	70	83.3	340000	8	AP001706
c	425	69.5	82.7	149233	14	AC119665
	426	69.5	82.7	283275	14	AC114372
c	427	69	82.1	311	4	BTU49780
c	428	69	82.1	641	10	BV061248
	429	69	82.1	1095	15	AK106177
	430	69	82.1	1105	15	AK104498
	431	69	82.1	1105	15	AK104803
	432	69	82.1	1233	15	AK068117
c	433	69	82.1	2845	15	SCYLR054C
	434	69	82.1	3029	2	BT021354
	435	69	82.1	3223	15	YSCSPT8TF
	436	69	82.1	3336	2	AF111942
c	437	69	82.1	3605	15	SCYLR055C
	438	69	82.1	3855	6	CQ614640
c	439	69	82.1	10907	6	CQ614639
	440	69	82.1	29143	2	AC115594
c	441	69	82.1	35913	15	AC149414

AC091545 Homo sapi
Continuation (6 of
BX890595 Zebrafish
AP000093 Homo sapi
AP000197 Homo sapi
CR352288 Zebrafish
Continuation (4 of
Continuation (5 of
Continuation (5 of
Continuation (3 of
Continuation (2 of
Continuation (256
CR383661 Zebrafish
AC135600 Oryza sat
AC032036 Homo sapi
AP002001 Homo sapi
AC133859 Oryza sat
AP000237 Homo sapi
AC018902 Homo sapi
AC113143 Homo sapi
AC164419 Mus muscu
AC044853 Homo sapi
AC148232 Didelphis
AP001985 Homo sapi
BX649370 Zebrafish
AC117186 Mus muscu
AC102446 Mus muscu
AC053477 Homo sapi
AC019041 Homo sapi
AC123713 Mus muscu
AC140214 Mus muscu
BX000486 Zebrafish
AC008074 Homo sapi
AC121818 Mus muscu
AC137110 Mus muscu
AC090977 Mus muscu
AC113022 Mus muscu
AC105714 Rattus no
AC123977 Mus muscu
AC130099 Rattus no
AC094779 Rattus no
AC099118 Rattus no
AC105519 Rattus no
AC109415 Rattus no
AC140433 Mus muscu
AC095568 Rattus no
AL807771 Mouse DNA
AC113874 Rattus no
AC102528 Mus muscu
AC108557 Rattus no
AC111654 Rattus no
AC095305 Rattus no
AC097144 Rattus no
AC093466 Mus muscu
AC126201 Rattus no
AP001706 Homo sapi
AC119665 Rattus no
AC114372 Rattus no
U49780 Bos taurus
BV061248 S212P6042
AK106177 Oryza sat
AK104498 Oryza sat
AK104803 Oryza sat
AK068117 Oryza sat
Z73226 S.cerevisia
BT021354 Drosophil
M94955 Saccharomyc
AF111942 Dictyoste
Z73227 S.cerevisia
CQ614640 Sequence
CQ614639 Sequence
AC115594 Dictyoste
AC149414 Phakopsor

	442	69	82.1	37983	14	CR385034
	443	69	82.1	40614	14	CR385033
c	444	69	82.1	40690	15	SCLACHXII
	445	69	82.1	41722	14	AC160846
	446	69	82.1	96482	14	AC136041
	447	69	82.1	110000	14	PFMAL8P1_10
	448	69	82.1	110000	15	CR382139_02
	449	69	82.1	110000	15	AP008207_101
	450	69	82.1	142268	15	AP003018
	451	69	82.1	148230	14	AC141548
	452	69	82.1	154248	15	AP003631
	453	69	82.1	154350	14	AC149951
c	454	69	82.1	160537	14	AC149948
c	455	69	82.1	161800	14	CR847528
	456	69	82.1	161905	9	AL929034
	457	69	82.1	161996	2	AC008284
c	458	69	82.1	163120	8	AC011454
	459	69	82.1	168062	14	AC012857
c	460	69	82.1	172479	2	AC006590
	461	69	82.1	176231	2	AC008367
c	462	69	82.1	181197	5	BX005330
	463	69	82.1	182211	8	AL360004
	464	69	82.1	182333	9	AC150285
	465	69	82.1	182883	9	AC134859
c	466	69	82.1	188649	14	AC161595
c	467	69	82.1	189786	9	AC140370
	468	69	82.1	198757	14	AC137192
c	469	69	82.1	201688	14	AC142194
c	470	69	82.1	203223	14	AC108997
c	471	69	82.1	203298	14	AC126806
	472	69	82.1	204657	9	AC162938
	473	69	82.1	205981	14	AC164881
	474	69	82.1	211279	5	BX120001
c	475	69	82.1	213581	9	AC147034
	476	69	82.1	216185	5	BX842696
	477	69	82.1	219213	9	AC154684
	478	69	82.1	224290	14	AC098199
c	479	69	82.1	239760	14	AC095113
c	480	69	82.1	240430	14	AC136419
	481	69	82.1	243184	14	AC097138
	482	69	82.1	251576	14	AC095068
c	483	69	82.1	252349	14	AC097200
c	484	69	82.1	252632	2	AE014818
c	485	69	82.1	253941	14	AC094334
c	486	69	82.1	256942	14	AC094745
c	487	69	82.1	260027	2	AE003659
c	488	69	82.1	260578	14	AC109104
c	489	69	82.1	261502	14	AC118131
c	490	69	82.1	266544	2	AC116956
	491	69	82.1	268305	14	AC097701
c	492	69	82.1	275031	14	CR936518
	493	69	82.1	302506	14	AC111719
c	494	69	82.1	304949	14	AC126717
c	495	69	82.1	307443	2	AE003751
c	496	69	82.1	328313	14	AC095391
c	497	69	82.1	339221	14	AC128827
c	498	69	82.1	347650	14	AC094577
	499	68.5	81.5	2094	6	CQ593436
c	500	68.5	81.5	4569	6	CQ593435
	501	68.5	81.5	45046	14	AC013093
c	502	68.5	81.5	170900	2	AC107487
c	503	68.5	81.5	178597	9	AL807824
c	504	68.5	81.5	198920	14	AC121393
	505	68.5	81.5	222874	14	AC105481
c	506	68.5	81.5	255672	14	AC123680
c	507	68.5	81.5	304419	2	AE003476
	508	68	81.0	409	2	PFA430087
	509	68	81.0	614	10	BV616252
c	510	68	81.0	628	10	BV277306
	511	68	81.0	1263	6	AR550357
	512	68	81.0	9003	2	AF411932
c	513	68	81.0	39575	2	AC116925
	514	68	81.0	69346	9	AL671997

CR385034 Homo sapi
CR385033 Homo sapi
X94607 S.cerevisia
AC160846 Homo sapi
AC136041 Rattus no
Continuation (11 o
Continuation (3 of
Continuation (102
AP003018 Oryza sat
AC141548 Rattus no
AP003631 Oryza sat
AC149951 Strongylo
AC149948 Strongylo
CR847528 Danio rer
AL929034 Mouse DNA
AC008284 Drosophil
AC011454 Homo sapi
AC012857 Drosophil
AC006590 Drosophil
AC008367 Drosophil
BX005330 Zebrafish
AL360004 Human DNA
AC150285 Mus muscu
AC134859 Mus muscu
AC161595 Mus muscu
AC140370 Mus muscu
AC137192 Rattus no
AC142194 Homo sapi
AC108997 Rattus no
AC126806 Mus muscu
AC162938 Mus muscu
AC164881 Mus muscu
BX120001 Zebrafish
AC147034 Mus muscu
BX842696 Zebrafish
AC154684 Mus muscu
AC098199 Rattus no
AC095113 Rattus no
AC136419 Rattus no
AC097138 Rattus no
AC095068 Rattus no
AC097200 Rattus no
AE014818 Plasmodiu
AC094334 Rattus no
AC094745 Rattus no
AE003659 Drosophil
AC109104 Rattus no
AC118131 Rattus no
AC116956 Dictyoste
AC097701 Rattus no
CR936518 Danio rer
AC111719 Rattus no
AC126717 Rattus no
AE003751 Drosophil
AC095391 Rattus no
AC128827 Rattus no
AC094577 Rattus no
CQ593436 Sequence
CQ593435 Sequence
AC013093 Drosophil
AC107487 Drosophil
AL807824 Mouse DNA
AC121393 Rattus no
AC105481 Rattus no
AC123680 Mus muscu
AE003476 Drosophil
AJ430087 Plasmodiu
BV616252 S215P6601
BV277306 S232P6158
AR550357 Sequence
AF411932 Plasmodiu
AC116925 Dictyoste
AL671997 Mouse DNA

c 515	68	81.0	110000	13	AY318871_1	Continuation (2 of
c 516	68	81.0	110000	14	AC118787_0	AC118787 Rattus no
c 517	68	81.0	110000	15	CR380957_09	Continuation (10 o
c 518	68	81.0	110000	15	AP008207_083	Continuation (84 o
c 519	68	81.0	114875	15	AC149430	AC149430 Populus t
c 520	68	81.0	143758	9	AC113265	AC113265 Mus muscu
c 521	68	81.0	161266	15	AP001278	AP001278 Oryza sat
522	68	81.0	163979	8	CNS01RGM	AL158800 Human chr
523	68	81.0	166677	9	AL590144	AL590144 Mouse DNA
c 524	68	81.0	167596	14	AC013177	AC013177 Drosophil
525	68	81.0	168556	5	BX842240	BX842240 Zebrafish
c 526	68	81.0	169546	14	AC004157	AC004157 Plasmodiu
c 527	68	81.0	176174	14	AC149918	AC149918 Strongylo
c 528	68	81.0	176977	14	AC105784	AC105784 Rattus no
c 529	68	81.0	183007	2	AC007257	AC007257 Drosophil
530	68	81.0	188078	9	AC129094	AC129094 Mus muscu
c 531	68	81.0	194449	9	AC134845	AC134845 Mus muscu
532	68	81.0	194959	14	AC155620	AC155620 Zea mays
533	68	81.0	196130	14	AC155492	AC155492 Zea mays
534	68	81.0	199590	9	AC133102	AC133102 Mus muscu
c 535	68	81.0	202803	9	AC093450	AC093450 Mus muscu
c 536	68	81.0	205800	5	BX072534	BX072534 Zebrafish
c 537	68	81.0	209986	9	AC124749	AC124749 Mus muscu
538	68	81.0	217285	14	AC145228	AC145228 Zea mays
539	68	81.0	224028	14	AC111276	AC111276 Rattus no
540	68	81.0	225001	14	AC126526	AC126526 Rattus no
c 541	68	81.0	230988	14	AC164388	AC164388 Bos tauru
542	68	81.0	231618	14	AC134749	AC134749 Rattus no
c 543	68	81.0	233933	14	AC128903	AC128903 Rattus no
544	68	81.0	245468	14	AC109534	AC109534 Rattus no
545	68	81.0	248047	14	AC135304	AC135304 Rattus no
546	68	81.0	250707	2	AE014848	AE014848 Plasmodiu
c 547	68	81.0	253541	14	AC108351	AC108351 Rattus no
548	68	81.0	269832	14	AC112751	AC112751 Rattus no
c 549	68	81.0	270456	14	AC111575	AC111575 Rattus no
550	68	81.0	271507	2	AE003625	AE003625 Drosophil
c 551	68	81.0	302156	2	AC116977	AC116977 Dictyoste
552	67.5	80.4	17986	2	U23451	U23451 Caenorhabdi
c 553	67.5	80.4	17986	2	U23451	U23451 Caenorhabdi
554	67.5	80.4	22683	2	AC006656	AC006656 Caenorhab
c 555	67.5	80.4	22683	2	AC006656	AC006656 Caenorhab
c 556	67.5	80.4	31361	2	AF016437	AF016437 Caenorhab
c 557	67.5	80.4	38094	2	AF040647	AF040647 Caenorhab
558	67.5	80.4	38383	2	CEF35H8	Z36752 Caenorhabdi
c 559	67.5	80.4	110000	15	AP008211_095	Continuation (96 o
c 560	67.5	80.4	150099	15	AC135916	AC135916 Genomic s
561	67.5	80.4	278007	14	AC006799	AC006799 Caenorhab
562	67	79.8	315	6	CQ646410	CQ646410 Sequence
563	67	79.8	318	6	AX608544	AX608544 Sequence
564	67	79.8	397	2	PFA430086	AJ430086 Plasmodiu
565	67	79.8	465	2	AJ698345	AJ698345 Euplotes
566	67	79.8	478	10	BV281117	BV281117 S232P6238
c 567	67	79.8	480	15	AY768786	AY768786 Passiflor
568	67	79.8	4779	6	CQ596381	CQ596381 Sequence
569	67	79.8	8979	2	AF411931	AF411931 Plasmodiu
570	67	79.8	9022	2	AF533700	AF533700 Plasmodiu
571	67	79.8	18724	2	AC115605	AC115605 Dictyoste
572	67	79.8	21693	14	AC017390	AC017390 Drosophil
c 573	67	79.8	22183	1	AE014286	AE014286 Streptoco
574	67	79.8	38449	15	AC152111	AC152111 Pichia st
575	67	79.8	39582	2	DMC17A9	AL009146 Drosophil
c 576	67	79.8	43539	6	AX602194	AX602194 Sequence
577	67	79.8	72006	8	AL607071	AL607071 Human DNA
578	67	79.8	81250	14	AC165439	AC165439 Medicago
c 579	67	79.8	103485	1	SAG766856	AL766856 Streptoco
c 580	67	79.8	110000	14	AC107201_4	Continuation (5 of
581	67	79.8	110000	15	AP008217_215	Continuation (216
c 582	67	79.8	110000	15	CR382131_38	Continuation (39 o
c 583	67	79.8	110000	15	CR382132_00	CR382132 Yarrowia
c 584	67	79.8	114189	14	AC157647	AC157647 Medicago
585	67	79.8	114252	5	BX649441	BX649441 Zebrafish
586	67	79.8	133130	14	AC141920	AC141920 Rattus no
c 587	67	79.8	136765	14	AC154937	AC154937 Loxodonta
c 588	67	79.8	136777	14	AC146986	AC146986 Strongylo
589	67	79.8	157881	14	CR936357	CR936357 Danio rer
c 590	67	79.8	168886	5	BX571698	BX571698 Zebrafish
591	67	79.8	169435	14	BX548039	BX548039 Danio rer
592	67	79.8	170681	15	AC146334	AC146334 Oryza sat
c 593	67	79.8	171424	9	AL831723	AL831723 Mouse DNA
594	67	79.8	177096	2	AC107326	AC107326 Drosophil
c 595	67	79.8	179053	5	CR376856	CR376856 Zebrafish
596	67	79.8	180365	9	AC154562	AC154562 Mus muscu
c 597	67	79.8	185568	14	AC145048	AC145048 Sus scrof
c 598	67	79.8	190641	9	AL626775	AL626775 Mouse DNA
599	67	79.8	192326	14	AC134317	AC134317 Rattus no
c 600	67	79.8	194318	9	AC150274	AC150274 Mus muscu

ALIGNMENTS

```
RESULT 1
AX594160
LOCUS       AX594160                  42 bp    DNA     linear   PAT 13-FEB-2003
DEFINITION  Sequence 13 from Patent WO02083910.
ACCESSION   AX594160
VERSION     AX594160.1  GI:28375390
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Farmer,A.A.
  TITLE     Sequence specific recombinase-based methods for producing intron
            containing vectors and compositions for use in practicing the same
  JOURNAL   Patent: WO 02083910-A 13 24-OCT-2002;
            Clontech Laboratories Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .42
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="splice sequence"
ORIGIN

Alignment Scores:
Pred. No.:                0.00102        Length:          42
Score:                    84.00          Matches:         12
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       6              Gaps:            0

US-09-858-332G-15 (1-12) x AX594160 (1-42)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db          4 CATAATCATAATCATAATCATAATCATAATCACAAC 39

RESULT 2
AX594158
LOCUS       AX594158                 100 bp    DNA     linear   PAT 13-FEB-2003
DEFINITION  Sequence 11 from Patent WO02083910.
ACCESSION   AX594158
VERSION     AX594158.1  GI:28375388
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Farmer,A.A.
  TITLE     Sequence specific recombinase-based methods for producing intron
            containing vectors and compositions for use in practicing the same
  JOURNAL   Patent: WO 02083910-A 11 24-OCT-2002;
            Clontech Laboratories Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .100
```

```
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="splice sequence"
ORIGIN

Alignment Scores:
Pred. No.:                0.00235        Length:          100
Score:                    84.00          Matches:         12
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       6              Gaps:            0

US-09-858-332G-15 (1-12) x AX594158 (1-100)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db         61 CATAATCATAATCATAATCATAATCATAATCACAAC 96

RESULT 3
AY559840/c
LOCUS       AY559840                 115 bp    DNA     linear   MAM 19-MAR-2004
DEFINITION  Bos taurus germline immunoglobulin heavy chain gene, DH3 segment,
            partial sequence.
ACCESSION   AY559840
VERSION     AY559840.1  GI:45551582
KEYWORDS    .
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 115)
  AUTHORS   Shojaei,F., Saini,S.S. and Kaushik,A.K.
  TITLE     Unusually long germline DH genes contribute to large sized CDR3H in
            bovine antibodies
  JOURNAL   Mol. Immunol. 40 (1), 61-67 (2003)
   PUBMED   12909131
REFERENCE   2  (bases 1 to 115)
  AUTHORS   Shojaei,F., Saini,S.S. and Kaushik,A.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-FEB-2004) Pathobiology, University of Guelph,
            University of Guelph, Guelph, ON N1G 2W1, Canada
FEATURES             Location/Qualifiers
     source          1. .115
                     /organism="Bos taurus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9913"
                     /chromosome="21"
                     /map="21q24"
                     /germline
     gene            <1. .>115
                     /gene="immunoglobulin heavy chain"
     D_segment       1. .115
                     /gene="immunoglobulin heavy chain"
                     /note="DH3"
     repeat_region   1. .9
                     /note="recombination signal sequence"
     repeat_region   23. .29
                     /note="recombination signal sequence"
     repeat_region   88. .94
                     /note="recombination signal sequence"
     repeat_region   107. .115
                     /note="recombination signal sequence"
ORIGIN

Alignment Scores:
Pred. No.:                0.00268        Length:          115
Score:                    84.00          Matches:         12
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       4              Gaps:            0

US-09-858-332G-15 (1-12) x AY559840 (1-115)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db         80 CATAACCATAACCATAACCACAACCATAACCATAAC 45

RESULT 4
AR568829/c
LOCUS       AR568829                 115 bp    DNA     linear   PAT 14-DEC-2004
DEFINITION  Sequence 3 from patent US 6740747.
ACCESSION   AR568829
VERSION     AR568829.1  GI:56568471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 115)
  AUTHORS   Kaushik,A. and Saini,S.S.
  TITLE     Bovine VDJ cassette, BF1H1, suitable for antigenization
  JOURNAL   Patent: US 6740747-A 3 25-MAY-2004;
FEATURES             Location/Qualifiers
     source          1. .115
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:                0.00268        Length:          115
Score:                    84.00          Matches:         12
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       6              Gaps:            0

US-09-858-332G-15 (1-12) x AR568829 (1-115)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db         80 CATAACCATAACCATAACCACAACCATAACCATAAC 45

RESULT 5
AF097207/c
LOCUS       AF097207                 387 bp    mRNA    linear   MAM 17-DEC-2003
DEFINITION  Bos taurus clone BF3H11 immunoglobulin heavy chain variable region
            mRNA, partial cds.
ACCESSION   AF097207
VERSION     AF097207.1  GI:3834654
KEYWORDS    .
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 387)
  AUTHORS   Saini,S.S. and Kaushik,A.
  TITLE     Extensive CDR3H length heterogeneity exists in bovine foetal VDJ
            rearrangements
  JOURNAL   Scand. J. Immunol. 55 (2), 140-148 (2002)
   PUBMED   11896930
REFERENCE   2  (bases 1 to 387)
  AUTHORS   Saini,S.S., Allore,B., Jacobs,R. and Kaushik,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1998) Pathobiology, University of Guelph, Guelph,
            ON N1G 2W1, Canada
FEATURES             Location/Qualifiers
     source          1. .387
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /clone="BF3H11"
                     /cell_type="mouse x Bovine heterohybridoma"
```

```
                     /dev_stage="fetal"
                     /note="PCR product"
     CDS             <1. .>387
                     /note="VDJ region"
                     /codon_start=1
                     /product="immunoglobulin heavy chain variable region"
                     /protein_id="AAC71042.1"
                     /db_xref="GI:3834655"
                     /translation="QVQLRESGPSLVKPSQTLSLTCTVSGFSLSSNGVGWVRQAPGKA
                     LEWVGGIDNDGDTYYNPALKSRLSITKDNSKSQVSLSVSSVTPEDTATYYCANSCYSG
                     YGYGCGYGYGYDYVDAWGQGLLVTVSS"
ORIGIN

Alignment Scores:
Pred. No.:                0.00864         Length:         387
Score:                    84.00           Matches:        12
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       4               Gaps:           0

US-09-858-332G-15 (1-12) x AF097207 (1-387)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        340 CATAACCATAACCATAACCACAACCATAACCATAAC 305

RESULT 6
AF097209/c
LOCUS       AF097209                390 bp    mRNA    linear   MAM 17-DEC-2003
DEFINITION  Bos taurus clone BF5F10 immunoglobulin heavy chain variable region
            mRNA, partial cds.
ACCESSION   AF097209
VERSION     AF097209.1  GI:3834658
KEYWORDS    .
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 390)
  AUTHORS   Saini,S.S. and Kaushik,A.
  TITLE     Extensive CDR3H length heterogeneity exists in bovine foetal VDJ
            rearrangements
  JOURNAL   Scand. J. Immunol. 55 (2), 140-148 (2002)
   PUBMED   11896930
REFERENCE   2  (bases 1 to 390)
  AUTHORS   Saini,S.S., Allore,B., Jacobs,R. and Kaushik,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1998) Pathobiology, University of Guelph, Guelph,
            ON N1G 2W1, Canada
FEATURES             Location/Qualifiers
     source          1. .390
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /clone="BF5F10"
                     /cell_type="mouse x Bovine heterohybridoma"
                     /dev_stage="fetal"
                     /note="PCR product"
     CDS             <1. .>390
                     /note="VDJ region"
                     /codon_start=1
                     /product="immunoglobulin heavy chain variable region"
                     /protein_id="AAC71044.1"
                     /db_xref="GI:3834659"
                     /translation="QVQLRESGPSLVKPSQTLSLTCTVSGFSLSDNSVGWVRQAPGKA
                     LEWLGVIYSGGSTGYNPALKSRLSITKDNSKSQVSLSLSSVTTEDTATYYCARDCYSG
                     YGYGYGCGYGYGSYVDAWGQGLLVTVSS"
ORIGIN

Alignment Scores:
Pred. No.:                0.0087          Length:         390
Score:                    84.00           Matches:        12
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       4               Gaps:           0

US-09-858-332G-15 (1-12) x AF097209 (1-390)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        340 CATAACCATAACCACAACCATAACCATAACCATAAC 305

RESULT 7
PFA430088
LOCUS       PFA430088               427 bp    DNA     linear   INV 07-FEB-2002
DEFINITION  Plasmodium falciparum partial rh1 gene, isolate FCB1.
ACCESSION   AJ430088
VERSION     AJ430088.1  GI:18642431
KEYWORDS    rh1 gene.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1
  AUTHORS   Taylor,H.M., Grainger,M. and Holder,A.A.
  TITLE     Variation in the expression of a Plasmodium falciparum protein
            family implicated in erythrocyte invasion
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 427)
  AUTHORS   Taylor,H.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-FEB-2002) Taylor H.M., Division of Parasitology,
            National Institute for Medical Research, Mill Hill, London, NW7
            1AA, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1. .427
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="FCB1"
                     /db_xref="taxon:5833"
                     /country="Colombia"
     gene            1. .426
                     /gene="rh1"
     CDS             <1. .>426
                     /gene="rh1"
                     /function="putative role in erythrocyte invasion"
                     /codon_start=1
                     /product="RH1 protein"
                     /protein_id="CAD23028.1"
                     /db_xref="GI:18642432"
                     /db_xref="UniProt/TrEMBL:Q8T8F1"
                     /translation="LENIKNNENIFDNIQLKKKDIDDIIININNTKETYLKELNKKKM
                     LQNKKKVDEKSEINNHHTLQHDNQNVEQKNKIKDHNLITKPNNNSSEESHQNEQMKEQ
                     NKNILEKQTRNIKPHHVHNHNHNHNHNHNHNHNHNQNQKD"
ORIGIN

Alignment Scores:
Pred. No.:                0.0095          Length:         427
Score:                    84.00           Matches:        12
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       2               Gaps:           0

US-09-858-332G-15 (1-12) x PFA430088 (1-427)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        358 CATAATCATAATCATAATCATAATCATAATCATAAT 393

RESULT 8
BV211484
LOCUS       BV211484                431 bp    DNA     linear   STS 04-JAN-2005
DEFINITION  BARC1076 Wheat var Chinese Spring Genomic DNA Triticum aestivum STS
```

```
            genomic, sequence tagged site.
ACCESSION   BV211484
VERSION     BV211484.1  GI:56847231
KEYWORDS    STS.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 431)
  AUTHORS   Song,Q.J., Shi,J.R., Singh,S., Fickus,E.W., Costa,J.M., Lewis,J.,
            Gill,B.S., Ward,R. and Cregan,P.B.
  TITLE     Development and Mapping of Microsatellite (SSR) Markers in Wheat
  JOURNAL   Theor. Appl. Genet. (2005) In press
COMMENT     Synonyms: XBARC1076
            Contact: Dr. Perry Cregan
            Soybean Genomics Improvement Laboratory
            United States Department of Agriculture
            BARC-WEST, Bldg. 006, Rm 100, 10300 Baltimore Ave., MD 20705, USA
            Tel: 301-504-5070
            Fax: 301-504-5728
            Email: creganp@ba.ars.usda.gov
            Primer A: CGCTGGTCACACTATGGGTCAAT
            Primer B: CGCATAAACCAACTTCTCTCATC
            STS size: 156
            PCR Profile:
                    Presoak:          95 degrees C for 3.0 minute(s)
                    Denaturation:     95 degrees C for 45 second(s)
                    Annealing:        48 degrees C for 45 second(s)
                    Polymerization:   72 degrees C for 45 second(s)
                    PCR Cycles:       35
                    Thermal Cycler: Perkin Elmer TC or MJ PTC-225 or MBS
            Satellite 384
            Protocol:
                    Template:         30-100 ng
                    Primer:           each 1.5 uM
                    dNTPs:            each 200 uM
                    Taq Polymerase:   0.05 units/ul
                    Total Vol:        10 ul

            Buffer:
                    MgCl2:            2.5 mM
                    KCl:              50 mM
                    Tris-HCl:         10 mM
                    pH:               8.3.
FEATURES             Location/Qualifiers
     source          1. .431
                     /organism="Triticum aestivum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4565"
                     /clone_lib="Wheat var Chinese Spring Genomic DNA"
                     /note="Vector: pBluescript; V-type: Plasmid; Genomic
                     fragments cloned into pBluescript"
     STS             <1. .>431
ORIGIN

Alignment Scores:
Pred. No.:                0.00958         Length:          431
Score:                    84.00           Matches:         12
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       10              Gaps:            0

US-09-858-332G-15 (1-12) x BV211484 (1-431)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        210 CATAATCATAATCATAATCATAATCATAATCATAAT 245

RESULT 9
BTU49779/c
LOCUS       BTU49779                  478 bp    mRNA    linear   MAM 16-AUG-2001
DEFINITION  Bos taurus immunoglobulin rearranged heavy chain variable region
            mRNA, partial cds.
ACCESSION   U49779
VERSION     U49779.1  GI:1293603
KEYWORDS    .
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 478)
  AUTHORS   Berens,S.J., Wylie,D.E. and Lopez,O.J.
  TITLE     Use of a single VH family and long CDR3s in the variable region of
            cattle Ig heavy chains
  JOURNAL   Int. Immunol. 9 (1), 189-199 (1997)
   PUBMED   9043960
REFERENCE   2  (bases 1 to 478)
  AUTHORS   Berens,S.J., Wylie,D. and Lopez,O.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-1996) Osvaldo J. Lopez, Biological Sciences,
            University of Nebraska, 325 Manter Hall, Lincoln, NE 68588-0118,
            USA
FEATURES             Location/Qualifiers
     source          1. .478
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /cell_type="heterohybridoma"
                     /tissue_type="spleen"
                     /dev_stage="150 day fetus"
     CDS             <1. .>478
                     /codon_start=2
                     /product="immunoglobulin heavy chain variable region"
                     /protein_id="AAA98660.1"
                     /db_xref="GI:1293604"
                     /translation="LLGVVTLICSKMNPPWTLLFVLSAPRGVLSQVQLRESGPSLVKP
                     SQTLSLTCTVSGFSLSDNSVGWVRQPPGKALEWLGVIYAGGSTGYNPALKSRLSITKD
                     NSKSQVSLSLSSVTTEDTATYYCARVGFCYSGYGYGYGCGYGYGDAWGQGLLVTVSS"
ORIGIN

Alignment Scores:
Pred. No.:                0.0106          Length:          478
Score:                    84.00           Matches:         12
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       4               Gaps:            0

US-09-858-332G-15 (1-12) x BTU49779 (1-478)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        437 CATAACCATAACCACAACCATAACCATAACCATAAC 402

RESULT 10
AF187065
LOCUS       AF187065                  519 bp    mRNA    linear   ROD 11-JUN-2000
DEFINITION  Rattus norvegicus p75NTR-associated cell death executor (Nade)
            mRNA, complete cds.
ACCESSION   AF187065
VERSION     AF187065.1  GI:8452895
KEYWORDS    .
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 519)
  AUTHORS   Mukai,J., Hachiya,T., Shoji-Hoshino,S., Kimura,M.T., Nadano,D.,
            Suvanto,P., Hanaoka,T., Li,Y., Irie,S., Greene,L.A. and Sato,T.A.
  TITLE     NADE, a p75NTR-associated cell death executor, is involved in
            signal transduction mediated by the common neurotrophin receptor
            p75NTR
```

```
  JOURNAL   J. Biol. Chem. 275 (23), 17566-17570 (2000)
   PUBMED   10764727
REFERENCE   2  (bases 1 to 519)
  AUTHORS   Mukai,J., Hachiya,T., Hoshino,S., Kimura,M., Nadano,D., Suvanto,P.,
            Hanaoka,T., Li,Y., Irie,S. and Sato,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-1999) Otolaryngology/Pathology, Columbia
            University, 630 West 168th St., P&S 11-451, New York, NY 10032, USA
FEATURES             Location/Qualifiers
     source          1. .519
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
     gene            1. .519
                     /gene="Nade"
     CDS             44. .406
                     /gene="Nade"
                     /note="involved in the common neurotrophin receptor
                     p75NTR-mediated signal transduction; NADE"
                     /codon_start=1
                     /product="p75NTR-associated cell death executor"
                     /protein_id="AAF75130.1"
                     /db_xref="GI:8452896"
                     /translation="MEQPLQNGQEDRPVGGGEGHQPAANNNNHNHNHNHNHNHNHNHH
                     RRGQARRLAPNFRWAIPNRQMNDGLGGDGDDMEMFMEEMREIRRKLRELQLRNCLRIL
                     MGELSNHHDHHDEFCLMP"
ORIGIN

Alignment Scores:
Pred. No.:              0.0115          Length:         519
Score:                  84.00           Matches:        12
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     9               Gaps:           0

US-09-858-332G-15 (1-12) x AF187065 (1-519)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        128 CACAACCATAACCACAACCACAACCACAACCACAAC 163

RESULT 11
BC058503
LOCUS       BC058503                887 bp    mRNA    linear   ROD 05-APR-2005
DEFINITION  Rattus norvegicus nerve growth factor receptor (TNFRSF16)
            associated protein 1, mRNA (cDNA clone MGC:72988 IMAGE:6887969),
            complete cds.
ACCESSION   BC058503
VERSION     BC058503.1  GI:34849868
KEYWORDS    MGC.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 887)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  CONSRTM   Mammalian Gene Collection Program Team
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
REFERENCE   2  (bases 1 to 887)
  AUTHORS   .
  CONSRTM   NIH MGC Project
  TITLE     Direct Submission
  JOURNAL   Submitted (16-SEP-2003) National Institutes of Health, Mammalian
            Gene Collection (MGC), Bethesda, MD 20892-2590, USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: John C. Marshall, M.D., Ph.D
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield,
            Andy Chan, Steve S. Chand, William Chow, Alison Cloutier, Ruth
            Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao,
            Kim MacDonald, Amara Masson, Mike R. Mayo, Josh Moran, Ryan Morin,
            Teika Olson, Diana Palmquist, Anca Petrescu, Anna Liisa Prahbu,
            Parvaneh Saeedi, JR Santos, Angelique Schnerch, Ursula Skalska,
            Duane Smailus, Jeff Stott, Miranda Tsai, George Yang, Jacquie
            Schein, Asim Siddiqui, Rob Holt, Marco Marra.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 52 Row: c Column: 1
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis, Similarity but not identity to protein.
FEATURES             Location/Qualifiers
     source          1. .887
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
                     /clone="MGC:72988 IMAGE:6887969"
                     /tissue_type="Pituitary gland, anterior, rat"
                     /clone_lib="NICHD_Rr_Pit1"
                     /lab_host="DH10B"
                     /note="Vector: pDNR-LIB"
     gene            1. .887
                     /gene="Ngfrap1"
                     /note="synonyms: Bex3, Nade"
                     /db_xref="GeneID:117089"
                     /db_xref="RGD:3148"
     CDS             218. .610
                     /gene="Ngfrap1"
                     /codon_start=1
                     /product="Ngfrap1 protein"
                     /protein_id="AAH58503.1"
                     /db_xref="GI:34849869"
                     /db_xref="GeneID:117089"
                     /db_xref="RGD:3148"
                     /translation="MANIHQENEEMEQPLQNGQEDRPVGGGEGHQPAANNNNHNHNHN
                     HNHNHNHNHHRRGQARRLAPNFRWAIPNRQMNDGLGGDGDDMEMFMEEMREIRRKLRE
                     LQLRNCLRILMGELSNHHDHHDEFCLMP"
ORIGIN

Alignment Scores:
Pred. No.:              0.0192          Length:         887
Score:                  84.00           Matches:        12
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     9               Gaps:           0
```

```
US-09-858-332G-15 (1-12) x BC058503 (1-887)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        332 CACAACCATAACCACAACCACAACCACAACCACAAC 367

RESULT 12
CNS06HW9/c
LOCUS       CNS06HW9                  994 bp    DNA     linear   STS 30-NOV-2001
DEFINITION  T3 end of clone AS0AA015G03 of library AS0AA from strain CLIB 533
            of Saccharomyces bayanus, sequence tagged site.
ACCESSION   AL399535
VERSION     AL399535.1  GI:12154132
KEYWORDS    STS.
SOURCE      Saccharomyces bayanus
  ORGANISM  Saccharomyces bayanus
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 994)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  PUBMED    11152876
REFERENCE   2  (bases 1 to 994)
  AUTHORS   Bon,E., Neuveglise,C., Casaregola,S., Artiguenave,F., Wincker,P.,
            Aigle,M. and Durrens,P.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 5.
            Saccharomyces bayanus var. uvarum
  JOURNAL   FEBS Lett. 487 (1), 37-41 (2000)
  PUBMED    11152880
REFERENCE   3  (bases 1 to 994)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This STS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .994
                     /organism="Saccharomyces bayanus"
                     /mol_type="genomic DNA"
                     /strain="CLIB 533"
                     /variety="uvarum"
                     /db_xref="taxon:4931"
                     /clone="AS0AA015G03"
                     /clone_lib="AS0AA"
                     /note="end : T3"
     STS             1. .994
ORIGIN

Alignment Scores:
Pred. No.:              0.0214          Length:         994
Score:                  84.00           Matches:        12
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     10              Gaps:           0
```

```
US-09-858-332G-15 (1-12) x CNS06HW9 (1-994)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        444 CATAACCATAACCATAACCATAACCACAACCACAAC 409

RESULT 13
AY065142/c
LOCUS       AY065142                  1132 bp    mRNA    linear   PLN 10-DEC-2001
DEFINITION  Arabidopsis thaliana putative protein (At4g21810; T8O5.20) mRNA,
            complete cds.
ACCESSION   AY065142
VERSION     AY065142.1  GI:17473758
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1132)
  AUTHORS   Nguyen,M., Karlin-Neumann,G., Southwick,A., Lam,B., Miranda,M.,
            Palm,C.J., Bowser,L., Jones,T., Banh,J., Carninci,P., Chen,H.,
            Cheuk,R., Chung,M.K., Hayashizaki,Y., Ishida,J., Kamiya,A.,
            Kawai,J., Kim,C., Lin,J., Liu,S.X., Narusaka,M., Pham,P.K.,
            Sakano,H., Sakurai,T., Satou,M., Seki,M., Shinn,P., Yamada,K.,
            Shinozaki,K., Ecker,J., Theologis,A. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-DEC-2001) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
COMMENT     e-mail for correspondence: arab@sequence.stanford.edu

            RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis  Full-Length cDNA'): Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members carried out the
            sequencing and annotation of the RAFL cDNAs: Nguyen,M.,
            Southwick,A., Karlin-Neumann,G.,  Lam,B., Miranda,M., Palm,C.J.,
            Bowser,L., Jones,T., Banh,J., Chen,H., Cheuk,R., Chung,M.K.,
            Kim,C., Lin,J., Liu,S.X., Pham,P.K., Sakano,H., Shinn,P.,
            Yamada,K., Ecker,J., Theologis,A. and Davis,R.W.

            Nguyen,M, (SSP/Stanford) and Seki,M. (RIKEN GSC) contributed
            equally to this work.  Shinozaki,K. (RIKEN GSC) and Davis,R.W.
            (SSP/Stanford)  contributed equally to this work as PIs.
FEATURES             Location/Qualifiers
     source          1. .1132
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /chromosome="IV"
                     /clone="RAFL06-09-I21"
                     /ecotype="Columbia"
                     /note="This clone is in pBluescript"
     gene            1. .1132
                     /gene="At4g21810; T8O5.20"
     CDS             150. .884
                     /gene="At4g21810; T8O5.20"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="AAL38318.1"
                     /db_xref="GI:17473759"
                     /translation="MAQAVEEWYKQMPIITRSYLTAAVVTTVGCSLEIISPYNLYLNP
                     TLVVKQYQFWRLVTNFLYFRKMDLDFLFHMFFLARYCKLLEENSFRGKTADFLYMLLF
                     GATVLTGIVLIGGMIPYLSVSFSKIIFLSNSLTFMMVYVWSKQNPYIHMSFLGLFTFT
                     AAYLPWVLLGFSILVGASAWGDFLGMIAGHAYYFLAFVYPRMTDRRPLKTPSFLKALF
                     ADEPVVIARPEDVRFAHAPFDEIHQD"
ORIGIN
```

```
Alignment Scores:
Pred. No.:                0.0243          Length:          1132
Score:                    84.00           Matches:         12
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       15              Gaps:            0

US-09-858-332G-15 (1-12) x AY065142 (1-1132)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        141 CACAATCACAATCACAATCACAATCACAATCACAAT 106

RESULT 14
AK107456
LOCUS       AK107456                1704 bp    mRNA    linear   PLN 24-JUL-2003
DEFINITION  Oryza sativa (japonica cultivar-group) cDNA clone:002-128-C10, full
            insert sequence.
ACCESSION   AK107456
VERSION     AK107456.1  GI:32992665
KEYWORDS    FLI_CDNA; oligo capping.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   The Rice Full-Length cDNA Consortium, National Institute of
            Agrobiological Sciences Rice Full-Length cDNA Project Team:,
            Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K.,
            Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I.,
            Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C.,
            Ohtsuki,K., Shishiki,T., Foundation of Advancement of International
            Science Genome Sequencing & Analysis Group:, Otomo,Y., Murakami,K.,
            Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y.,
            Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M.,
            Narikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Niikura,J.,
            Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Miura,J.,
            Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN:,
            Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuda,S.,
            Hara,A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M.,
            Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Osato,N., Ota,Y.,
            Saito,R., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
            Yoshino,M. and Hayashizaki,Y.
  TITLE     Collection, mapping, and annotation of over 28,000 cDNA clones from
            japonica rice
  JOURNAL   Science 301 (5631), 376-379 (2003)
   PUBMED   12869764
REFERENCE   2  (bases 1 to 1704)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K.,
            Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiraoka,T.,
            Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K.,
            Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I.,
            Kanagawa,S., Katoh,H., Kawagashira,N., Kawai,J., Kawamata,M.,
            Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M.,
            Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Kurosaki,T., Kusumegi,T., Li,C., Lu,M.,
            Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A.,
            Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakamura,M.,
            Namiki,T., Narikawa,R., Niikura,J., Nishi,K., Nomura,K.,
            Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M., Ooka,H.,
            Osato,N., Ota,Y., Otomo,Y., Ryu,R., Saitoh,H., Sakai,C., Sakai,K.,
            Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Shibata,K.,
            Shinagawa,A., Shiraki,T., Shishiki,T., Sogabe,Y., Sugano,S.,
            Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y.,
            Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A.,
            Toya,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,W.,
            Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and
            Yoshimura,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of
            Agrobiological Sciences, Department of Molecular Genetics, Head of
            Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki
            305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp,
            Tel:81-29-838-7007, Fax:81-29-838-7007)
COMMENT     This clone is one of the 28K full-length cDNA clones from japonica
            rice.
            URL : http://cdna01.dna.affrc.go.jp/cDNA/
            NIAS Rice Full-Length cDNA Project Team: Kikuchi,S., Satoh,K.,
            Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J.,
            Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T.,
            Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T. and
            Yamamoto,M.
            FAIS Genome Sequencing & Analysis Group: Otomo,Y., Iida,Y.,
            Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M., Kobayashi,M.,
            Kodama,T., Kurosaki,T., Kusumegi,T., Lu,M., Masuda,H., Miura,J.,
            Mizuno,K., Narikawa,R., Niikura,J., Oka,M., Ryu,R., Sugano,S.,
            Sugiyama,A., Suzuki,Y., Tsunoda,Y., Ueda,M., Xie,Q., Yokomizo,S.,
            Yoshimura,A., Matsubara,K. and Murakami,K.
            Genome Exploration Research Group in Riken Genomic Sciences Center
            and Genome Science Laboratory in Riken: Adachi,J., Aizawa,K.,
            Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hanagaki,T.,
            Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K.,
            Hiraoka,T., Hori,F., Iida,J., Imamura,K., Imotani,K., Ishii,Y.,
            Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J.,
            Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Osato,N.,
            Ota,Y., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H.,
            Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F.,
            Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K.,
            Yasunishi,A. and Hayashizaki,Y.
FEATURES              Location/Qualifiers
     source           1. .1704
                      /organism="Oryza sativa (japonica cultivar-group)"
                      /mol_type="mRNA"
                      /cultivar="Nipponbare"
                      /db_xref="taxon:39947"
                      /clone="002-128-C10"
ORIGIN

Alignment Scores:
Pred. No.:                0.036           Length:          1704
Score:                    84.00           Matches:         12
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       15              Gaps:            0

US-09-858-332G-15 (1-12) x AK107456 (1-1704)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        307 CACAACCACAACCACAACCACAATCACAACCACAAT 342

RESULT 15
I34914
LOCUS       I34914                  1897 bp    DNA     linear   PAT 13-MAY-1997
DEFINITION  Sequence 1 from patent US 5599701.
ACCESSION   I34914
VERSION     I34914.1  GI:2087882
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1897)
  AUTHORS   Henry,S.A. and White,M.J.
  TITLE     Inositol-excreting yeast
  JOURNAL   Patent: US 5599701-A 1 04-FEB-1997;
FEATURES              Location/Qualifiers
     source           1. .1897
                      /organism="unknown"
                      /mol_type="unassigned DNA"
```

```
ORIGIN

Alignment Scores:
Pred. No.:                0.0399          Length:          1897
Score:                    84.00           Matches:         12
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       6               Gaps:            0

US-09-858-332G-15 (1-12) x I34914 (1-1897)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db       1730 CATAATCATAATCATAATCATAATCATAATCATAAT 1765

RESULT 16
YSCOPI1
LOCUS       YSCOPI1                 1897 bp    DNA     linear   PLN 27-APR-1993
DEFINITION  Saccharomyces cerevisiae negative phospholipid biosynthesis
            regulator (OPI1) gene, complete cds.
ACCESSION   M57383 J05727
VERSION     M57383.1  GI:172067
KEYWORDS    negative phospholipid biosynthesis regulator.
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 1897)
  AUTHORS   White,M.J., Hirsch,J.P. and Henry,S.A.
  TITLE     The OPI1 gene of Saccharomyces cerevisiae, a negative regulator of
            phospholipid biosynthesis, encodes a protein containing
            polyglutamine tracts and a leucine zipper
  JOURNAL   J. Biol. Chem. 266 (2), 863-872 (1991)
   PUBMED   1985968
COMMENT     Original source text: S.cerevisiae DNA.
FEATURES             Location/Qualifiers
     source          1. .1897
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA" .
                     /db_xref="taxon:4932"
     gene            439. .1653
                     /gene="OPI1"
     CDS             439. .1653
                     /gene="OPI1"
                     /codon_start=1
                     /product="negative phospholipid biosynthesis regulator"
                     /protein_id="AAA34828.1"
                     /db_xref="GI:172068"
                     /translation="MSENQRLGLSEEEVEAAEVLGVLKQSCRQKSQPSEDVSQADKMP
                     ASESSTTPLNILDRVSNKIISNVVTFYDEINTNKRPLKSIGRLLDDDDDEHDDYDYND
                     DEFFTNKRQKLSRAIAKGKDNLKEYKLNMSIESKKRLVTCLHLLKLANKQLSDKISCL
                     QDLVEKEQVHPLHKQDGNARTTTGAGEDETSSDEDDDDEEFFDASEQVNASEQSIVVK
                     MEVVGTVKKVYSLISKFTANSLPEPARSQVRESLLNLPTNWFDSVHSTSLPHHASFHY
                     ANCEEQKVEQQQQQQQQQQQQLLQQQLLQQQQQKRNKDGDDSASPSSSVTANGKVLI
                     LAKESLEMVRNVMGVVDSTLGKAEEWVKQKQEVKEMIRERFLQQQQQYRQQQQKDGNY
                     VKPSQDNVDSKD"
ORIGIN

Alignment Scores:
Pred. No.:                0.0399          Length:          1897
Score:                    84.00           Matches:         12
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       15              Gaps:            0

US-09-858-332G-15 (1-12) x YSCOPI1 (1-1897)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db       1730 CATAATCATAATCATAATCATAATCATAATCATAAT 1765

RESULT 17
AX536592
LOCUS       AX536592                1898 bp    DNA     linear   PAT 22-NOV-2002
DEFINITION  Sequence 193 from Patent WO02064766.
ACCESSION   AX536592
VERSION     AX536592.1  GI:25262997
KEYWORDS    .
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1
  AUTHORS   Contreras,R.H., Eberhardt,I., Luyten,W.H. and Reekmans,R.J.
  TITLE     Bax-responsive genes for drug target identification in yeast and
            fungi
  JOURNAL   Patent: WO 02064766-A 193 22-AUG-2002;
            JANSSEN PHARMACEUTICA N.V. (BE)
FEATURES             Location/Qualifiers
     source          1. .1898
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:4932"
ORIGIN

Alignment Scores:
Pred. No.:                0.04            Length:          1898
Score:                    84.00           Matches:         12
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       6               Gaps:            0

US-09-858-332G-15 (1-12) x AX536592 (1-1898)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        195 CATAATCATAATCATAATCATAATCATAATCATAAT 230

RESULT 18
BC029741
LOCUS       BC029741                1949 bp    mRNA    linear   ROD 07-OCT-2003
DEFINITION  Mus musculus deleted in polyposis 1-like 1, mRNA (cDNA clone
            MGC:25862 IMAGE:4195889), complete cds.
ACCESSION   BC029741
VERSION     BC029741.1  GI:20988223
KEYWORDS    MGC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1949)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
```

```
   PUBMED   12477932
REFERENCE   2  (bases 1 to 1949)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
            Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
            A.N., Gibbs, R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 30 Row: p Column: 24
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 21314829.
FEATURES             Location/Qualifiers
     source          1. .1949
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="MGC:25862 IMAGE:4195889"
                     /tissue_type="Liver, normal. 5 month old male mouse."
                     /clone_lib="NCI_CGAP_Li9"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     gene            1. .1949
                     /gene="Dpl11"
                     /db_xref="GeneID:70335"
                     /db_xref="MGI:1917585"
     CDS             74. .598
                     /gene="Dpl11"
                     /codon_start=1
                     /product="Dpl11 protein"
                     /protein_id="AAH29741.1"
                     /db_xref="GI:20988224"
                     /db_xref="GeneID:70335"
                     /db_xref="MGI:1917585"
                     /translation="MDGLRQRFERFLEQKNVATEALGALEARTGVEKRYLAAGALALL
                     GLYLLFGYGASLLCNVIGFVYPAYASVKAIESPSKEDDTVWLTYWVVYALFGLVEFFS
                     DLLLFWFPFYYAGKCAFLLFCMTPGPWNGALLLYHRVIRPLFLKHHMALDSAASQLSG
                     RALDLAAGITRDAK"
     misc_feature    245. .505
                     /gene="Dpl11"
                     /note="TB2_DP1_HVA22; Region: TB2/DP1, HVA22 family. This
                     family includes members from a wide variety of eukaryotes.
                     It includes the TB2/DP1 (deleted in polyposis) protein,
                     which in humans is deleted in severe forms of familial
                     adenomatous polyposis, an autosomal dominant oncological
                     inherited disease. The family also includes the plant
                     protein of known similarity to TB2/DP1, the HVA22 abscisic
                     acid-induced protein, which is thought to be a regulatory
                     protein"
                     /db_xref="CDD:pfam03134"
ORIGIN

Alignment Scores:
Pred. No.:              0.041          Length:        1949
Score:                  84.00          Matches:       12
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     9              Gaps:          0

US-09-858-332G-15 (1-12) x BC029741 (1-1949)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        776 CACAACCACAACCACAACCACAACCACAACCACAAC 811

RESULT 19
AB039933
LOCUS       AB039933               2011 bp    mRNA    linear   ROD 25-FEB-2005
DEFINITION  Mus musculus Dpl11 mRNA for polyposis locus protein 1-like 1 (TB2
            protein-like 1), complete cds.
ACCESSION   AB039933
VERSION     AB039933.1  GI:7573634
KEYWORDS    polyposis locus protein 1-like 1.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Sato,H., Tomita,H., Nakazawa,T., Wakana,S. and Tamai,M.
  TITLE     Deleted in Polyposis 1-like 1 Gene (Dpl11): A Novel Gene Richly
            Expressed in Retinal Ganglion Cells
  JOURNAL   Invest. Ophthalmol. Vis. Sci. 46, 791-796 (2005)
REFERENCE   2  (bases 1 to 2011)
  AUTHORS   Sato,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-2000) Hajime Sato, Tohoku University School of
            Medicine, Department of Ophthalmology; Seiryo-machi 1-1, Sendai,
            Miyagi, 980-8574, Japan (E-mail:hasato@oph.med.tohoku.ac.jp,
            Tel:81-22-717-7294, Fax:81-22-717-7298)
FEATURES             Location/Qualifiers
     source          1. .2011
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL6"
                     /db_xref="taxon:10090"
                     /chromosome="10"
                     /clone="pSK138"
                     /tissue_type="retina"
                     /dev_stage="adult"
                     /note="vector:pBluescriptSK(-)"
     gene            1. .2011
                     /gene="Dpl11"
     CDS             62. .667
                     /gene="Dpl11"
                     /standard_name="deleted in polyposis 1-like 1"
                     /codon_start=1
                     /product="polyposis locus protein 1-like 1 (TB2
                     protein-like 1)"
                     /protein_id="BAA94544.1"
                     /db_xref="GI:7573635"
                     /translation="MDGLRQRFERFLEQKNVATEALGALEARTGVEKRYLAAGALALL
                     GLYLLFGYGASLLCNVIGFVYPAYASVKAIESPSKEDDTVWLTYWVVYALFGLVEFFS
                     DLLLFWFPFYYAGKCAFLLFCMTPGPWNGALLLYHRVIRPLFLKHHMALDSAASQLSG
                     RALDLAAGITRDVLQALARGRALVTPASTSEPPAALELDPK"
     polyA_signal    1968. .1973
                     /gene="Dpl11"
ORIGIN

Alignment Scores:
Pred. No.:              0.0422         Length:        2011
Score:                  84.00          Matches:       12
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     9              Gaps:          0

US-09-858-332G-15 (1-12) x AB039933 (1-2011)
```

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        845 CACAACCACAACCACAACCACAACCACAACCACAAC 880

RESULT 20
BC083830
LOCUS       BC083830                2152 bp    mRNA    linear   ROD 30-MAR-2005
DEFINITION  Rattus norvegicus deleted in polyposis 1-like 1 (predicted), mRNA
            (cDNA clone MGC:94922 IMAGE:7112359), complete cds.
ACCESSION   BC083830
VERSION     BC083830.1  GI:54035486
KEYWORDS    MGC.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 2152)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  CONSRTM   Mammalian Gene Collection Program Team
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2152)
  AUTHORS   .
  CONSRTM   NIH MGC Project
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) National Institutes of Health, Mammalian
            Gene Collection (MGC), Bethesda, MD 20892-2590, USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Howard Jacobs
            cDNA Library Preparation: Express Genomics
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 182 Row: h Column: 17
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Similarity but not
            identity to protein.
FEATURES             Location/Qualifiers
     source          1. .2152
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
                     /clone="MGC:94922 IMAGE:7112359"
                     /tissue_type="Testis, rat (Brown Norway)"
                     /clone_lib="NIH_MGC_237"
                     /lab_host="DH10B"
                     /note="Vector: pExpress1"
     gene            1. .2152
                     /gene="Dpl11_predicted"
                     /db_xref="GeneID:362835"
                     /db_xref="RGD:1309508"
     CDS             28. .663
                     /gene="Dpl11_predicted"
                     /codon_start=1
                     /product="deleted in polyposis 1-like 1 (predicted)"
                     /protein_id="AAH83830.1"
                     /db_xref="GI:54035487"
                     /db_xref="GeneID:362835"
                     /db_xref="RGD:1309508"
                     /translation="MDGLRQRFERFLEQKNVATDALGALEARTGVEKRYLAAGALTLL
                     GLYLLFGYGASLLCNVIGFVYPAYASVKAIESPNKEDDTVWLTYWVVYALFGLVEFFS
                     DLLLFWFPFYYAGKCAFLLFCMTPGPWNGALLLYHRVIRPLFLKHHVALDSAASQLSG
                     RALDIAAGITRDVLQALARGRTLVTPASASESPAALEPDPKSSQTTLLKHK"
ORIGIN

Alignment Scores:
Pred. No.:                0.0451          Length:          2152
Score:                    84.00           Matches:         12
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       9               Gaps:            0

US-09-858-332G-15 (1-12) x BC083830 (1-2152)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        919 CACAACCACAACCACAACCACAACCACAACCACAAC 954

RESULT 21
AX461526/c
LOCUS       AX461526                2270 bp    DNA     linear   PAT 08-JUL-2002
DEFINITION  Sequence 455 from Patent WO0198480.
ACCESSION   AX461526
VERSION     AX461526.1  GI:21726737
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Budworth,P., Brown,D., Chang,H.S., Zhu,T., Han,B., Wang,X. and
            Cooper,B.
  TITLE     Promoters for regulation of plant gene expression
  JOURNAL   Patent: WO 0198480-A 455 27-DEC-2001;
            Syngenta Participations AG (CH)
FEATURES             Location/Qualifiers
     source          1. .2270
                     /organism="Arabidopsis thaliana"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:3702"
ORIGIN

Alignment Scores:
Pred. No.:                0.0475          Length:          2270
Score:                    84.00           Matches:         12
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       6               Gaps:            0

US-09-858-332G-15 (1-12) x AX461526 (1-2270)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
```

```
Db        604 CACAATCACAATCACAATCACAATCACAATCACAAT 569

RESULT 22
AF020407
LOCUS       AF020407               2396 bp    DNA     linear   INV 11-MAY-1998
DEFINITION  Dictyostelium discoideum WimA (wimA) gene, partial cds.
ACCESSION   AF020407 AF012944
VERSION     AF020407.1  GI:2425142
KEYWORDS    .
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 2396)
  AUTHORS   Loomis,W.F. and Iranfar,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-AUG-1997) Biology, UCSD, 9500 Gilman Drive, La Jolla,
            CA 92093, USA
COMMENT     On May 11, 1998 this sequence version replaced gi:2393950.
FEATURES             Location/Qualifiers
     source          1. .2396
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
     gene            702. .>2396
                     /gene="wimA"
                     /note="disruption results in morphological defect;
                     developmental gene"
     CDS             702. .>2396
                     /gene="wimA"
                     /codon_start=1
                     /product="WimA"
                     /protein_id="AAC16025.1"
                     /db_xref="GI:2425143"
                     /translation="MENQDSSLFFGEPDPQIDFFDSLSSGQEQQEQQQEQQQEQQQEQ
                     QEQTQPQTQQEQPTSHFEDGESLFESLGQQDSNSTNTTTNIESPPHQQDDQSFFDSFN
                     SSDSGVGITNEQPKTEQPTQPFSNLIQPNEVIELQSDVIQHQNQVIEHQNSIIEQHQR
                     HQHQFSSDSSNIGIVDSLIENPILVSDNVVNSLSDSIVGKDDDETSPNNNNIDNIDNI
                     DNIDNIDDNNNNTNNNNNSGNQENENLSTSSSNEHTPAQSEPIVTTTEGSNNNENNI
                     INNNENISRDDDNSNSNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNSGEKEK
                     EISSFFSSDDNVNNSEVLFFNSTNSDTNSSANSIETVSVTSAIAAEEKDASSFFNNIS
                     ASQELQPQPQPQPQPQPQPQPQPQPQPQQQTQQQQSQTQQSPITNKPPVVEKDISSFF
                     ATPEMESQNDPSGNIFQSFNISSNTPKMVSQVPQPQPQQQTQPPPPQQQQQPQQPLQT
                     PTKSQPQQTEKDVSSFFGGGSSSFKQPQQPERDVSSFFGGGSTTNEVSSLFGSMTID"
ORIGIN

Alignment Scores:
Pred. No.:              0.05          Length:        2396
Score:                  84.00         Matches:       12
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     2             Gaps:          0

US-09-858-332G-15 (1-12) x AF020407 (1-2396)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db       1897 CACAACCACAACCACAACCACAACCACAACCACAAC 1932

RESULT 23
AX594148
LOCUS       AX594148               4938 bp    DNA     linear   PAT 13-FEB-2003
DEFINITION  Sequence 1 from Patent WO02083910.
ACCESSION   AX594148
VERSION     AX594148.1  GI:28375378
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Farmer,A.A.
  TITLE     Sequence specific recombinase-based methods for producing intron
            containing vectors and compositions for use in practicing the same
  JOURNAL   Patent: WO 02083910-A 1 24-OCT-2002;
            Clontech Laboratories Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .4938
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Synthetic Vector"
ORIGIN

Alignment Scores:
Pred. No.:              0.1           Length:        4938
Score:                  84.00         Matches:       12
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-15 (1-12) x AX594148 (1-4938)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        141 CATAATCATAATCATAATCATAATCATAATCACAAC 176

RESULT 24
AX594150
LOCUS       AX594150               6525 bp    DNA     linear   PAT 13-FEB-2003
DEFINITION  Sequence 3 from Patent WO02083910.
ACCESSION   AX594150
VERSION     AX594150.1  GI:28375380
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Farmer,A.A.
  TITLE     Sequence specific recombinase-based methods for producing intron
            containing vectors and compositions for use in practicing the same
  JOURNAL   Patent: WO 02083910-A 3 24-OCT-2002;
            Clontech Laboratories Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .6525
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Synthetic Vector"
ORIGIN

Alignment Scores:
Pred. No.:              0.131         Length:        6525
Score:                  84.00         Matches:       12
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-15 (1-12) x AX594150 (1-6525)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db       1728 CATAATCATAATCATAATCATAATCATAATCACAAC 1763

RESULT 25
AX594151
LOCUS       AX594151               7487 bp    DNA     linear   PAT 13-FEB-2003
DEFINITION  Sequence 4 from Patent WO02083910.
ACCESSION   AX594151
VERSION     AX594151.1  GI:28375381
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
```

```
REFERENCE   1
  AUTHORS   Farmer,A.A.
  TITLE     Sequence specific recombinase-based methods for producing intron
            containing vectors and compositions for use in practicing the same
  JOURNAL   Patent: WO 02083910-A 4 24-OCT-2002;
            Clontech Laboratories Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .7487
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Synthetic Vector"
ORIGIN

Alignment Scores:
Pred. No.:                0.15          Length:          7487
Score:                    84.00         Matches:         12
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       6             Gaps:            0

US-09-858-332G-15 (1-12) x AX594151 (1-7487)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db       2316 CATAATCATAATCATAATCATAATCATAATCACAAC 2351

RESULT 26
AY160101
LOCUS       AY160101                7801 bp    DNA     linear   INV 23-OCT-2002
DEFINITION  Dictyostelium discoideum nucleotide exchange factor RasGEF L (gefL)
            gene, complete cds.
ACCESSION   AY160101
VERSION     AY160101.1  GI:24286698
KEYWORDS    .
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 7801)
  AUTHORS   Wilkins,A., Szafranski,K., Gloeckner,G., Harrisingh,M.,
            Deenadayalan,B., Mueller,R., Eichinger,L., Noegel,A.A. and
            Insall,R.
  TITLE     The family of rasGEF genes in Dictyostelium discoideum
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 7801)
  AUTHORS   Szafranski,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-2002) Abt. Genomanalyse, Institut fuer Molekulare
            Biotechnologie, Beutenbergstrasse 11, D-07745 Jena, Germany
FEATURES             Location/Qualifiers
     source          1. .7801
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
     gene            <296. .>5023
                     /gene="gefL"
     mRNA            <296. .>5023
                     /gene="gefL"
                     /product="nucleotide exchange factor RasGEF L"
     CDS             296. .5023
                     /gene="gefL"
                     /note="GefL"
                     /codon_start=1
                     /product="nucleotide exchange factor RasGEF L"
                     /protein_id="AAN46881.1"
                     /db_xref="GI:24286699"
                     /translation="MGSPIYKSFQSLYETLLDDLDYSKKRREELLNLSVEEKWIFLQR
                     IKPDCTNSLVSNGVYQLQHQISSNDLKTLISSPPKLNSSQQLLSSSSSSSSLSSSSTS
                     ATHSLQNKSLPSISVCPKSPRTHMKVLSASSLFTLEDITKLRQQLVSKPVSWVINFAE
                     STGIVSILSMITTYSQNLINNVNKDQPQSQPQPQPQPQPRPQPQQPQPQQPQPQQLLP
                     LLSQLQPQQQQQQQQQTQQPQPQLQAQTKSDDHFIINECLGCIKLLVELEMQSLLLND
                     TAEIIFPLILSSQVSTKRMTFEVLELICKIFSLGSTIVLESIKNYSKKINEPLTKVFS
                     HITSPLDSEYPADLKVSCMSLINHLIGSNHLSLEARIETRSKFLSNDILLIIKNLRID
                     GEGIETSSWKQLELEMKQFESMMVQDDNQTSNAEANELVSNELSQSIDNNNNNNNNNN
                     NNSSTNLVNNNNNNINSNCNSIINGGNSSDSLINNSVGTGVGSIMGVGSFMGVNEIE
                     YDSPRKEFDELKSSNNNNSNSNSVNNFINNITEPIMPLTPTEGKLVVIVSMQQTRQQA
                     QLQQQLQQQLEEEDDDEVPLSSSVATTASTTANTTLIEPNHLPIQTREVIVSVQRKSQ
                     AGDVIRSLLASNSNLKELGEWGLFVNDSNSNNNNNNNNNNNNKLSTNVIGRFLKDDEY
                     IFDQMTTSSTSGSLSSDDGATPRLMTRSIPRFKEYRLKMIPWKVRVSLEKVKEMGLVL
                     NANNTANGQFNESKFIAEAPMDPTMTCTGLVYHLIKKYLPNVNDRYCLEADDFGLYLD
                     TNAGGLMWGGGTGGGYWLEPLEKLHDYEIFKDPKSTVALKLRPKGVKLKFADGTYEIF
                     RLDLMSPTEKIFQEISEKVIDSNNNINLSYYGIFVDRQQGGANGNGISDSPILVSQSS
                     STSLIEESKKQRKPEWLEKGLPLYHYKINSRCCLRFALRPSNITLLIDISLFKSFQQQ
                     QQQQQQQQQQSTSSSSSSSSSSVSSNNNSTNNSNNLIGSLNNSINTISELDESIDDN
                     GLSSSSSDKEDQISTPSTSNLLSKFQQVDESISNLKFQLPQPIQSKQIIVQLPLHLPL
                     GEALETDENWIETGIEREKCSFHFDSFDGILMNQTQPLNTQNFTEKNRICVLMADELE
                     INSEQTNHNKTNIWDEPFDSSTIICDPDNPGQIRAATLNKLLEQATNNIEKDRELMNI
                     LLMTYMSFATSDILLDKLVERYNVPEHEKEQKSVVQLHTIVFLKNWLEQQAPQAASGG
                     GLEEKFLVRLSEFVDKQISLDGYTKIVPQLKKWIDSALKEKRAYAMPEAARPNQLTTK
                     LPSLTSSLLDDELFVAQQLTLREYETFKRIQAVEFLGQAWGKPKLQYKAPNLLKMIER
                     FNKISTGVSTAILSQPKIKTRVKLICRFIKIAQHCRDLNNYHLLTAFMAGIRNSNVVR
                     LRLSWAKVPKKHLHILEDLEKIMSMEGSFKEFRTKMAETIPPCIPYLGVYLKDLTFIE
                     EGNPDSINSLINWGKKKLIYNIISIIQRCQQVSYDFGNQNQSQVAQKSDLVLNYFDSL
                     PIALDELLYKISLNLEPRSNFNN"
ORIGIN

Alignment Scores:
Pred. No.:                0.156         Length:          7801
Score:                    84.00         Matches:         12
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       2             Gaps:            0

US-09-858-332G-15 (1-12) x AY160101 (1-7801)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db        855 CACAATCACAACCACAACCACAACCACAACCACAAC 890

RESULT 27
AF411930
LOCUS       AF411930                9024 bp    DNA     linear   INV 04-AUG-2003
DEFINITION  Plasmodium falciparum strain HB3 normocyte-binding protein 1 (NBP1)
            gene, complete cds.
ACCESSION   AF411930
VERSION     AF411930.2  GI:33414602
KEYWORDS    .
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 9024)
  AUTHORS   Rayner,J.C., Vargas-Serrato,E., Huber,C.S., Galinski,M.R. and
            Barnwell,J.W.
  TITLE     A Plasmodium falciparum Homologue of Plasmodium vivax Reticulocyte
            Binding Protein (PvRBP1) Defines a Trypsin-resistant Erythrocyte
            Invasion Pathway
  JOURNAL   J. Exp. Med. 194 (11), 1571-1582 (2001)
   PUBMED   11733572
REFERENCE   2  (bases 1 to 9024)
  AUTHORS   Rayner,J.C., Barnwell,J.W., Galinski,M.R., Tran,T., Huber,C.S. and
            Corredor,V.
  TITLE     Differing geographic diversity among members of the Reticulocyte
            Binding Like (RBL) gene family of the two major human malaria
            parasites, Plasmodium falciparum and Plasmodium vivax
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 9024)
  AUTHORS   Rayner,J.C., Galinski,M.R. and Barnwell,J.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-AUG-2001) Division of Parasitic Diseases, Centers for
            Disease Control and Prevention, 4770 Buford Highway, MS F-13,
            Chamblee, GA 30341, USA
REFERENCE   4  (bases 1 to 9024)
  AUTHORS   Rayner,J.C., Barnwell,J.W. and Galinski,M.R.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-2002) Division of Parasitic Diseases, Centers for
            Disease Control and Prevention, 4770 Buford Highway, MS F-13,
            Chamblee, GA 30341, USA
  REMARK    Sequence update by submitter
COMMENT     On Aug 4, 2003 this sequence version replaced gi:17352157.
FEATURES             Location/Qualifiers
     source          1. .9024
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /strain="HB3"
                     /db_xref="taxon:5833"
     gene            <8. .>9024
                     /gene="NBP1"
     mRNA            join(<8. .62,170. .>9024)
                     /gene="NBP1"
                     /product="normocyte-binding protein 1"
     CDS             join(8. .62,170. .9024)
                     /gene="NBP1"
                     /note="similar to Plasmodium vivax reticulocyte binding
                     protein 1"
                     /codon_start=1
                     /product="normocyte-binding protein 1"
                     /protein_id="AAL38219.2"
                     /db_xref="GI:33414603"
                     /translation="MQRWIFCNIVLHILIYLAEFSHEQESYSSNEKIRKDYSDDNNYE
                     PTPSYEKRKKEYGKDESYIKNYRGNNFSYDLSKNSSIFLHMGNGSNSKTLKRCNKKKN
                     IKTNFLRPIEEEKTVLNNYVYKGVNFLDTIKRNDSSYKFDVYKDTSFLKNREYKELIT
                     MQYDYAYLEATKEVLYLIPKDKDYHKFYKNKLEKILFNLKDSLKLLREGYIQSKLEMI
                     RIHSDIDILNEFHQGNIINDNYFNNEIKKKKEDMEKYIREYNLYIYKYENQLKIKIQK
                     LTNEVSINLNKSTCEKNCYNYILKLEKYKNIIKDKINKWKDLPEIYIDDKSFSYTFLK
                     DVINNKIDIYKTISSFISTQKQLYYLEYIYIMNKNTLNLLSYNIQKTDINSSSKYTYT
                     KSHFLKDNHILLSKYYTAKFIDILNKTYYYNLYKNKILLFNKYIIKLRNDLKEYAFKS
                     IQFIQDKIKKHKDELSIENILQEVNNIYIKYDTSINEISKYNNLIINTDLQIVQQKLL
                     EIKQKKNDITHKVQLINHIYKNIHDEILNKKNNEITKIIINNIKDHKKDLQDLLLFIQ
                     QIKQYNILTDHNITQCNNYYKEIIKMKEDINHIHIYIQPILNNLHTLKQVQNNKIKYE
                     EHIKQILQKIYDKKESLKKIILLKDEAQLDITLLDDLIQKQTKKQTQKQTQTQKQTLI
                     QNNETIQLISGQEDKHESNPFNHIQTYIQQKDTQNKNIQNLLKSLYNGNINTFIDTIS
                     KYILKQKDIELTQHVYTDEKINDYLEEIKNEQNKIDKTIDDIKIQETLKQITHIVNNI
                     KTIKKDLLKEFIQHLIKYMNERYQNMQQGYNNLTNYINQYEEENNNMKQYITTIRNIQ
                     KIYYDNIYAKEKEIRSGQYYKDFITSRKNIYNIRENISKNVDMIKNEEKKKIQNCVDK
                     YNSIKQYVKMFKNGDTQDENNNNNNDIYDKLIVPLDSIKQNIDKYNTEHNFITFTNKI
                     NTHNKKNQEMMEEFIYAYKRLKILKILNISLKACEKNNKSINTLNDKTQELKKIVTHE
                     IDLLQKDILTSQISNKNVLLLNDLLKEIEQYIIDVHKLKKKSNDLFTYYEQSKNYFYF
                     KNKKDNFDIQKTINKMNEWLAIKNYINEINKNYQTLYEKKINVLLHNSKSYVQYFYDH
                     IINLILQKKNYLENTLKTKIQDNEHSLYALQQNEEYQKVKNEKDQNEIKKIKQLIEKN
                     KNDILTYENNIEQIEQKNIELKTNAQNKDDQIVNTLNEVKKKIIYTYEKVDNQISNVL
                     KNYEEGKVEYDKNVVQNVNDADDTNDIDEINDIDEINDIDEIKDIDHIKHFDDTKHFD
                     DIYHADDTRDEYHIALSNYIKTELRNINLQEIKNNIIKIFKEFKSAHKEIKKESEQIN
                     KEFTKMDVVINQLRDIDRQMLDLYKELDEKYSEFNKTKIEEINNIRENINNVEIWYEK
                     NIIEYFLRHMNDQKDKAAKYMENIDTYKNNIEIISKQINPENYVETLNKSNMYSYVEK
                     ANDLFYKQINNIIINSNQLKNEAFTIDELQNIQKNRKNLLTKKQQIIQYTNEIENIFN
                     EIKNINNILVLTNYKSILQDISQNINHVSIYTEQLHNLYIKLEEEKEQMKTLYHKSNV
                     LHNQINFNEDAFINNLLINIEKIKNDITHIKEKTNIYMIDVNKSKNNAQLYFHNTLRG
                     NEKIEYLKNLKNSTNQQITLQELKQVQENVEKVKDIYNQTIKYEEEIKKNYHIITDYE
                     NKINDILHNSFIKQINMESSNNKKQTKQIIDIINDKTFEEHIKTSKTKINMLKEQSQM
                     KHIDKTLLNEQALKLFVDINSTNNNLDNMLSEINSIQNNIHTYIQEANKSFDKFKIIC
                     DQNVNDLLNKLSLGDLNYMNHLKNLQNEIRNMNLEKNFMLDKSKKIDEEEKKLDILKV
                     NISNINNSLDKLKKYYEEALFQKVKEKAEIQKENIEKIKQEINTLSDVFKKPFFFIQL
                     NTDSSQHEKDINNNVETYKNNIDEIYNVFIQSYNLIQKYSSEIFSSTLNYIQTKEIKE
                     KSIKEQNQLNQNEKEASVLLKNIKINETIKLFKQIKNERQNDVHNIKEDYNLLQQYLN
                     YMKNEMEQLKKYKIDVHMDKNYVENNNGEKEKLLKETISSYYDKINNINNKLYIYKNK
                     EDTYFNNMIKVSEILNIIIKKKQQNEQRIVINAEYDSSLINKDEEIKKEINNQIIELN
                     KHNENISNIFKDIQNIKKQSQDIITNMNDMCKSTILLVDIIQKKEEALNKQKNILRNI
                     DNILNKKENIIDKVIKCNCDDYKDILIQNETEYQKLQNINHTYEEKKKSIDILKIKNI
                     KQKNIQEYKNKLEQMNTIINQSIEQHVFINADILQNEKIKLEEIIKNLDILDEQIMTY
                     HNSIDELYKLGIQCDNHLITTISVVVNKNTTKIMIHIKKQKEDIQKINNYIQTNYNII
                     NEEALQFHRLYGHNLISEDDKNNLVHIIKEQKNIYTQKEIDISKIIKHVKKGLYSLNE
                     HDMNHDTHMNIINEHINNNILQPYTQLINMIKDIDNVFIKIQNNKFEQIQKYIEIIKS
                     LEQLNKNINTDNLNKLKDTQNKLINIETEMKHKQKQLINKMNDIEKDNITDQYMHDVQ
                     QNIFEPITLKMNEYNTLLNDDHNNNINNEHQFNHLNSLHTKIFSHNYNKEQQQEYITN
                     IMQRIDVFINDLDTYQYEYYFYEWNQEYKQIDKNKINQHINNIKNNLIHVKKQFEHTL
                     ENIKNNENIFDNIQLKKKDIDDIIININNTKETYLKELNKKKMLQNKKKVDEKSEINN
                     HHTLQHDNQNVEQKNKIKDHNLITKPNNNSSEESHQNEQMKEQNKNILEKQTRNIKPH
                     HVHNHNHNHNHNHNHNHNHNQNQKDSTKLQEQDISTHKLHNTIHEQQSKDNHQGNREK
                     KQKNGNHERMYFASGIVVSILFLSSLGFVINSKNNKQEYDKEQEKQQQNDFVCDNNKM
                     DDKSTQKYGRNQEEVMEISFDNDYI"
ORIGIN

Alignment Scores:
Pred. No.:                0.179          Length:          9024
Score:                    84.00          Matches:         12
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       2              Gaps:            0

US-09-858-332G-15 (1-12) x AF411930 (1-9024)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db       8605 CATAATCATAATCATAATCATAATCATAATCATAAT 8640

RESULT 28
AF411933
LOCUS       AF411933                9044 bp    DNA     linear   INV 04-AUG-2003
DEFINITION  Plasmodium falciparum strain 7G8 normocyte-binding protein 1 (NBP1)
            gene, complete cds.
ACCESSION   AF411933
VERSION     AF411933.2  GI:33414608
KEYWORDS    .
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 9044)
  AUTHORS   Rayner,J.C., Vargas-Serrato,E., Huber,C.S., Galinski,M.R. and
            Barnwell,J.W.
  TITLE     A Plasmodium falciparum Homologue of Plasmodium vivax Reticulocyte
            Binding Protein (PvRBP1) Defines a Trypsin-resistant Erythrocyte
            Invasion Pathway
  JOURNAL   J. Exp. Med. 194 (11), 1571-1582 (2001)
   PUBMED   11733572
REFERENCE   2  (bases 1 to 9044)
  AUTHORS   Rayner,J.C., Barnwell,J.W., Galinski,M.R., Tran,T., Huber,C.S. and
            Corredor,V.
  TITLE     Differing geographic diversity among members of the Reticulocyte
            Binding Like (RBL) gene family of the two major human malaria
            parasites, Plasmodium falciparum and Plasmodium vivax
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 9044)
  AUTHORS   Rayner,J.C., Galinski,M.R. and Barnwell,J.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-AUG-2001) Division of Parasitic Diseases, Centers for
            Disease Control and Prevention, 4770 Buford Highway, MS F-13,
            Chamblee, GA 30341, USA
REFERENCE   4  (bases 1 to 9044)
  AUTHORS   Rayner,J.C., Barnwell,J.W. and Galinski,M.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-2002) Division of Parasitic Diseases, Centers for
            Disease Control and Prevention, 4770 Buford Highway, MS F-13,
            Chamblee, GA 30341, USA
  REMARK    Sequence update by submitter
COMMENT     On Aug 4, 2003 this sequence version replaced gi:17352163.
FEATURES             Location/Qualifiers
     source          1. .9044
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /strain="7G8"
                     /db_xref="taxon:5833"
     gene            <8. .>9026
                     /gene="NBP1"
     mRNA            join(<8. .62,151. .>9026)
                     /gene="NBP1"
                     /product="normocyte-binding protein 1"
     CDS             join(8. .62,151. .9026)
                     /gene="NBP1"
```

/note="similar to Plasmodium vivax reticulocyte binding protein 1"
/codon_start=1
/product="normocyte-binding protein 1"
/protein_id="AAL38222.2"
/db_xref="GI:33414609"
/translation="MQRWIFCNIVLHILIYLAEFSHEQESYSSNEKIRKDYSDDNNYE
PTPSYEKRKKEYGKDESYIKNYRGNNFSYDLSKNSSIFLHMGNGSNSKTLKRCNKKKN
IKTNFLRPIEEEKTVLNNYVYKGVNFLDTIKRNDSSYKFDVYKDTSFLKNREYKELIT
MQYDYAYLEATKEVLYLIPKDKDYHKFYKNELEEILFNLKDSLKLLREGYIQSKLEMI
RIHSDIDILNEFHQGNIINDNYFNNEIKKKKEDVEKYIREYNLYIYKYENQLKIKIQK
LTNEVSINLNKSTCEKNCYNYILKLEKYKNIIKDKINKWKDLPEIYIDDKSFSYTFLK
DVINNKIDIYKTISSFISTQKQLYYLEYIYIMNKNTLNLLSYNIQKTDINSSSKYTYT
KSHFLKDNHILLSKYYTAKFIDILNKTYYYNLYKNKILLFNKYIINLRNDLKEYAFKS
IQFIQDKIKKHKDELSIENILQEVNNIYIKYDTSINEISKYNNLIINTDLQIVQQKLL
EIKQKKNDITHKVQLINHIYKNIHDEILNKKNNEITKIIINNIKDHKKDLQDLLLFIQ
QIKQYNILTDHNITQCNNYYKEIIKMKEDINHIHIYIQPILNNLHTLKQVQNNKIKYE
EHIKQILQKIYDKKESLKKIILLKDEAQLDITLLDDLIQKQTKKQTQTQTQTQKQTLI
QNNETIQLISGQEDKHESNPFNHIQTYIQQKDTQNKNIQNLLKSLYNGNINTFIDTIS
KYILKQKDIELTQHVYTDEKINDYLEEIKNEQNKIDKTIDDIKIQETLKQITHIVNNI
KTIKKDLLKEFIQHLIKYMNERYQNMQQGYNNLTNYINQYEEENNNMKQYITTIRNIQ
KIYYDNIYAKEKEIRSGQYYKDFITSRKNIYNIRENISKNVDMIKNEEKKKIQNCVDK
YNSIKQCVKMLKNGDTQDENNNNNNNNNNDIYDKLIVPLDSIKQNIDKYNTEHNFIT
FTNKINTHNKKNQEMMEEFIYAYKRLKILKILNISLKACEKNNKSINTLNDKTQELKK
IVTHEIDLLQKDILTSQISNKNVLLLNDLLKEIEQYIIDVHKLKKKSNDLFTYYEQSK
NYFYFKNKKDNFDIQKTINKMNEWLAIKNYINEINKNYQTLYEKKINVLLHNSKSYVQ
YFYDHIINLILQKKNYLENTLKTKIQDNEHSLYALQQNEEYQKVKNEKDQNEIKKIKQ
LIEKNKNDILTYENNIEQIEQKNIELKTNAQNKDDQIVNTLNEVKKKIIYTYEKVDNQ
ISNVLKNYEEGKVEYDKNVVQNVNDADDTNDIDEINDIDEINDIDEIKDIDH
IKHFDDTKHFDDIYHADDTRDEYHIALSNYIKTELRNINLQEIKNNIIKIFKEFKSAH
KEIKKESKQINKEFTKMDVVINQLRDIDRQMLDLYKELDEKYSEFNKTKIEEINNIRE
NINNVEIWYEKNIIEYFLRHMNDQKDKAAKYMENIDTYKNNIEIISKQINPENYVETL
NKSNMYSYVEKANDLFYKQINNIIINSNQLKNEAFTIDELQNIQKNRKNLLTKKQQII
QYTNEIENIFNEIKNINNILVLTNYKSILQDISQNINHVSIYTEQLHNLYIKLEEEKE
QMKTLYHKSNVLHNQINFNEDAFINNLLINIEKIKNDITHIKEKTNIYMIDVNKSKNN
AQLYFHNTLRGNEKIEYLKNLKNSTNQQITLQELKQVQENVEKVKDIYNQTIKYEEKI
KKNYHIITDYENKINDILHNSFIKQINMESSNNKKQTKQIIDIINDKTFEEHIKTSKT
KINMLKEQSQMKHIDKTLLNEQALKLFVDINSTNNNLDNMLSEINSIQNNIHTYIQEA
NKSFDKFKIICDQNVNDLLNKLSLGDLNYMNHLKNLQNEIRNMNLEKNFMLDKSKKID
EEEKKLDILKVNISNINNSLDKLKKYYEEALFQKVKEKAENQKENIEKIKQEINTLSD
VFKKPFFFIQLNTDSSQHEKDINNNVETYKNNIDEIYNVFIQSYNLIQKYSSEIFSST
LNYIQTKEIKEKSIKEQNQLNQNEKEASVLLKNIKINETIKLFKQIKNERQNDVHNIK
EDYNLLQQYLNYMKNEMEQLKKYKNDVHMDKNYVENNNGEKEKLLKETISSYYDKINN
INNKLYIYKNKEDTYFNNMIKVSEILNIIIKKKQQNEQRIVINAEYDSSLINKDEEIK
KEINNQIIELNKHNENISNIFKDIQNIKKQSQDIITNMNDMHKSTILLVDIIQKKEEA
LNKQKNILRNIDNILNKKENIIDKVIKCNCDDYKDILIQNETEYQKLQNINHTYEEKK
KSIDILKIKNIKQKNIQEYKNKLEQMNTIINQSIEQHVFINADILQNEKIKLEEIIKN
LDILDEQIMTYHNSIDELYKLGIQCDNHLITTISVVVNKNTTKIMIHIKKQKEDIQKI
NNYIQTNYNIINEEALQFHRLYGHNLISEDDKNNLVHIIKEQKNIYTQKEIDISKIIK
HVKKGLYSLNEHDMNHDTHMNIINEHINNNILQPYTQLINMIKDIDNVFIKIQNNKFE
QIQKYIEIIKSLEQLNKNINTDNLNKLKDTQNKLINIETEMKHKQKQLINKMNDIEKD
NITDQYMHDVQQNIFEPITLKMNEYNTLLNDNHNNNINNEHQFNHLNSLHTKIFSHNY
NKEQQQEYITNIMQRIDVFINDLDTYQYEYYFYEWNQEYKQIDKNKINQHINNIKNNL
IHVKKQFEHTLENIKNNENIFDNIQLKKKDIDDIIINININNTKETYLKELNKKIKMLQNK
KKVDEKSEINNHHTLQHDNQNVEQKNKIKDHNLITKPNNNSSEESHQNEQMKEQNKNI
LEKQTRNIKPHHVHNHNHNHNHNHNQNQNQKDSTKLQEQDISTHKLHNTIHEQQSKDN
HQGNREKKQKNGNHERMYFASGIVVSILFLSSLGFVINSKNNKQEYDKEQEKQQQNDF
VCDNNKMDDKSTQKYGRNQEEVMEISFDNDYI"

ORIGIN

Alignment Scores:

Pred. No.:	0.18	Length:	9044
Score:	84.00	Matches:	12
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-15 (1-12) x AF411933 (1-9044)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db       8619 CATAATCATAATCATAATCATAATCATAATCATAAT 8654
```

RESULT 29
AF411929
LOCUS AF411929 9049 bp DNA linear INV 04-AUG-2003
DEFINITION Plasmodium falciparum strain FVO normocyte-binding protein 1 (NBP1) gene, complete cds.
ACCESSION AF411929
VERSION AF411929.2 GI:33414600
KEYWORDS .
SOURCE Plasmodium falciparum (malaria parasite P. falciparum)
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 9049)
AUTHORS Rayner,J.C., Vargas-Serrato,E., Huber,C.S., Galinski,M.R. and Barnwell,J.W.
TITLE A Plasmodium falciparum Homologue of Plasmodium vivax Reticulocyte Binding Protein (PvRBP1) Defines a Trypsin-resistant Erythrocyte Invasion Pathway
JOURNAL J. Exp. Med. 194 (11), 1571-1582 (2001)
PUBMED 11733572
REFERENCE 2 (bases 1 to 9049)
AUTHORS Rayner,J.C., Barnwell,J.W., Galinski,M.R., Tran,T., Huber,C.S. and Corredor,V.
TITLE Differing geographic diversity among members of the Reticulocyte Binding Like (RBL) gene family of the two major human malaria parasites, Plasmodium falciparum and Plasmodium vivax
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 9049)
AUTHORS Rayner,J.C., Barnwell,J.W. and Galinski,M.R.
TITLE Direct Submission
JOURNAL Submitted (22-AUG-2001) Division of Parasitic Diseases, Centers for Disease Control and Prevention, 4770 Buford Highway, MS F-13, Chamblee, GA 30341, USA
REFERENCE 4 (bases 1 to 9049)
AUTHORS Rayner,J.C., Barnwell,J.W. and Galinski,M.R.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-2002) Division of Parasitic Diseases, Centers for Disease Control and Prevention, 4770 Buford Highway, MS F-13, Chamblee, GA 30341, USA
REMARK Sequence update by submitter
COMMENT On Aug 4, 2003 this sequence version replaced gi:17352155.
FEATURES Location/Qualifiers
source 1. .9049
/organism="Plasmodium falciparum"
/mol_type="genomic DNA"
/strain="FVO"
/db_xref="taxon:5833"
gene <8. .>9048
/gene="NBP1"
mRNA join(<8. .62,170. .>9048)
/gene="NBP1"
/product="normocyte-binding protein 1"
CDS join(8. .62,170. .9048)
/gene="NBP1"
/note="similar to Plasmodium vivax reticulocyte binding protein 1"
/codon_start=1
/product="normocyte-binding protein 1"
/protein_id="AAL38218.2"
/db_xref="GI:33414601"
/translation="MQRWIFCNIVLHILIYLAEFSHEQESYSSNEKIRKDYSDDNNYE
PTPSYEKRKKEYGKDESYIKNYRGNNFSYDLSKNSSIFLHMGNGSNSKTLKRCNKKKN
IKTNFLRPIEEEKTVLNNYVYKGVNFLDTIKRNDSSYKFDVYKDTSFLKNREYKELIT
MQYDYAYLEATKEVLYLIPKDKDYHKFYKNKLEKILFNLKDSLKLLREGYIQSKLEMI
RIHSDIDILNEFHQGNIINDNYFNNEIKKKKEDMEKYIREYNLYIYKYENQLKIEIQK
LTNEVSINLNKSTCEKNCYYYILKLEKYKNIIKDKINKWKDLPEIYIDDKSFSYTFLK
DVINNKIDIYKTISSFISTQKQLYYLEYIYIMNKNTLNLLSYNIQKTDINSSSKYTYT
KSHFLKDNHILLSKYYTAKFIDILNKTYYYNLYKNKILLFNKYIIKLRNDLKEYAFKS
IQFIQDKIKKHKDELSIENILQEVNNIYIKYDTSINEISKYNNLIINTDLQIVQQKLL
EIKQKKNDITHKVQLINHIYKNIHDEILNKKNNEITKIIINNIKDHKKDLQDLLLFIQ
QIKQYNILTDHNITQCNNYYKEIIKMKEDINHIHIYIQPILNNLHTLKQVQNNKIKYE
EHIKQILQKIYDKKESLKKIILLKDEAQLDITLLDDLIQKQTQTQTQTQKQTLIQNNE
TIQLISGQEDKHESNPFNHIQTYIQQKDTQNKNIQNLLKSLYNGNINTFIDTISKYIL

KQKDIELTQHVYTDEKINDYLEEIKNEQNKIDKTIDDIKIQETLKQITHIVNNIKTIK
KDLLKEFIQHLIKYMNERYQNMQQGYNNLTNYINQYEEENNNMKQYITTIRNIQKIYY
DNIYAKEKEIRSGQYYKDFITSRKNIYNIRENISKNVDMIKNEEKKKIQNCVDKYNSI
KQYVKMFKNGDTQDENNNNNNDIYDKLIVPLDSIKQNIDKYNTEHNFITFTNKINTHN
KKNQEMMEEFIYVYKRLKILKILNISLKACEKNNKSINTLNDKTQELKKIVTHEIDLL
QKDILTSQISNKNVLLLNDLLKEIEQYIIDVHKLKKKSNDLFTYYEQSKNYFYFKNKK
DNFDIQKTINKMNEWLAIKNYINEINKNYQTLYEKKINVLLHNSKSYVQYFYDHIINL
ILQKKNYLENTLKTKIQDNEHSLYALQQNEEYQKVKNEKDQNEIKKIKQLIEKNKNDI
LTYENNIEQIEQKNIELKTNAQNKDDQIVNTLNEVKKKIIYTYEKVDNQISNVLKNYE
EGKVEYDKNVVQNVNDADDTNDIDEINDIDEINDIDEINDIDEIKDIDHIKH
FDDTKHFDDIYHADDTRDEYHIALSNYIKTELRNINLQEIKNNIIKIFKEFESAHKEI
KKESKQINKEFTKMDVVINQLRDIDRQMLDLYKELDEKYSEFNKTKIEEINNIRENIN
NVEIWYEKNIIEYFLRHMNDQKDKAAKYMENIDTYKNNIEIISKQINPENYVETLNKS
NMYSYVEKANDLFYKQINNIIINSNQLKNEAFTIDELQNIQKNRKNLLTKKQQIIQYT
NEIENIFNEIKNINNILVLTNYKSILQDISQNINHVSIYTEQLHNLYIKLEEEKEQMK
TLYHKSNVLHNQINFNEDAFINNLLINIEKIKNDITHIKEKTNIYMIDVNKSKNNAQL
YFHNTLRGNEKIEYLKNLKNSTNQQITLQELKQVQENVEKVKDIYNQTIKYEEKIKKN
YHIITDYENKINDILHNSFIKQINMESSNNKKQTKQIIDIINDKTFEEHIKTSKTKIN
MLKEQSQMKHIDKTLLNEQALKLFVDINSTNNNLDNMLSEINSIQNNIHTYIQEANKS
FDKFKIICDQNVNDLLNKLSLGDLNYMNHLKNLQNEIRNMNLEKNFMLDKSKKIDEEE
KKLDILKVNISNINNSLDKLKKYYEEALFQKVKEKAENQKENIEKIKQEINTLSDVFK
KPFFFIQLNTDSSQHEKDINNNVETYKNNIDEIYNVFIQSYNLIQKYSSEIFSSTLNY
IQTKEIKEKSIKEQNQLNQNEKEAFVLLKNIKINETIKLFKQIKNERQNDVHNIKEDY
NLLQQYLNYMKNEMEQLKKYKNDVHMDKNYVENNNGEKEKLLKETISSYYDKINNINN
KLYIYKNKEDTYFNNMIKVSEILNIIIKKKQQNEQRIVINAEYDSSLINKDEEIKKEI
NNQIIELNKHNENISNIFKDIQNIKKQSQDIITNMNDMHKSTILLVDIIQKKEEALNK
QKNILRNIDNILNKKENIIDKVIKCNCDDYKDILIQNETEYQKLQNINHTYEEKKKSI
DILKIKNIKQKNIQEYKNKLEQMNTIINQSIEQHVFINADILQNEKIKLEEIIKNLDI
LDEQIMTYHNSIDELYKLGIQCDNHLITTISVVVNKNTTKIMIHIKKQKEDIQKINNY
IQTNYNIINEEALQFHRLYGHNLISEDDKNNLVHIIKEQKNIYTQKEIDISKIIKHVK
KGLYSLNEHDMNHDTHMNIINEHINNNILQPYTQLINMIKDIDNVFIKIQNNKFEQIQ
KYIEIIKSLEQLNKNINTDNLNKLKDTQNKLINIETEMKHKQKQLINKMNDIEKDNIT
DQYMHDVQQNIFEPITLKMNEYNTLLNDDHNNNINNEHQFNHLNSLHTKIFSHNYNKE
QQQEYITNIMQRIDVFINDLDTYQYEYYFYEWNQEYKQIDKNKINQHINNIKNNLIHV
KKQFEHTLENIKNNENIFDNIQLKKKDIDDIIINININNTKETYLKELNKKKMLQNKKKV
DEKSEINNHHTLQHDNQNVEQKNKIKDHNLITKPNNNSSEESHQNEQMKEQNKNILEK
QTRNIKPHHVHNHNHNHNHNHNHNHNHNQNQKDSTKLQEQDISTHKLHNTIHEQQSKD
NHQGNREKKQKNGNHERMYFASGIVVSILFLSSLGFVINSKNNKQEYDKEQEKQQQND
FVCDNNKMDDKSTQKYGRNQEEVMEISFDNDYI"

ORIGIN

Alignment Scores:

Pred. No.:	0.18	Length:	9049
Score:	84.00	Matches:	12
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-15 (1-12) x AF411929 (1-9049)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db       8629 CATAATCATAATCATAATCATAATCATAATCATAAT 8664
```

RESULT 30
AY520825

LOCUS	AY520825 20247 bp DNA linear ROD 01-MAR-2005
DEFINITION	Mus musculus ISP1 (Isp1) and ISP2 (Isp2) genes, complete cds.
ACCESSION	AY520825
VERSION	AY520825.1 GI:49487500
KEYWORDS	.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 20247)
AUTHORS	O'Sullivan,C.M., Tang,L., Xu,H. and Rancourt,D.E.
TITLE	Origin of the Implanation Serin Proteinase Subfamily in Mouse
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 20247)
AUTHORS	O'Sullivan,C.M., Tang,L., Xu,H. and Rancourt,D.E.
TITLE	Direct Submission
JOURNAL	Submitted (08-JAN-2004) Biochemistry & Molecular Biology, U of Calgary, 3330 Hospital Dr. N.W., Calgary, AB T2J 1N4, Canada

FEATURES	Location/Qualifiers
source	1. .20247 /organism="Mus musculus" /mol_type="genomic DNA" /strain="129/SvJ" /db_xref="taxon:10090" /chromosome="17" /map="A3.3"
gene	4730. .7968 /gene="Isp1"
mRNA	join(4730. .4891,5496. .5556,5738. .5918,6020. .6285, 7321. .7484,7610. .7968) /gene="Isp1" /product="ISP1"
5'UTR	4730. .4891 /gene="Isp1"
CDS	join(5496. .5556,5738. .5918,6020. .6285,7321. .7484, 7610. .7762) /gene="Isp1" /note="serine proteinase" /codon_start=1 /product="ISP1" /protein_id="AAT66743.1" /db_xref="GI:49487501" /translation="MFRLLLLALSCLESTVFMASVSISRSKPVGIVGGQRTPPGKWPW QVSLRMYSYEVNSWVHICGGSIIHPQWILTAAHCIQSQDADPAVYRVQVGEVYLYKEQ ELLNISRIIIHPDYNDVSKRFDLALMQLTALLVTSTNVSPVSLPKDSSTFDSTDQCWL VGWGNLLQRVPLQPPYQLHEVKIPIQDNKSCKRAYRKKSSDEHKAVAIFDDMLCAGTS GRGPCFGDSGGPLVCWKSNKWIQVGVVSKGIDCSNNLPSIFSRVQSSLAWIHQHIQ"
3'UTR	7763. .7968 /gene="Isp1"
misc_feature	7969. .14673 /note="intergenic region"
gene	14674. .18578 /gene="Isp2"
mRNA	join(14674. .14836,16305. .16365,16518. .16698,16894. .17159, 17990. .18156,18241. .18578) /gene="Isp2" /product="ISP2"
5'UTR	14674. .14836 /gene="Isp2"
CDS	join(16305. .16365,16518. .16698,16894. .17159,17990. .18156, 18241. .18405) /gene="Isp2" /note="serine proteinase" /codon_start=1 /product="ISP2" /protein_id="AAT66744.1" /db_xref="GI:49487502" /translation="MLIQLCLTLFFLGCSIAGTPAPGPEDVLMGIVGGHSAPQGKWPW QVSLRIYRYYWAFWVHNCGGSIIHPQWVLTAAHCIRERDADPSVFRIRVGEAYLYGGK ELLSVSRVIIHPDFVHAGLGSDVALLQLAVSVQSFPNVKPVKLPSESLEVTKKDVCWV TGWGAVSTHRSLPPPYRLQQVQVKIIDNSLCEEMYHNATRHRNRGQKLILKDMLCAGN QGQDSCYGDSGGPLVCNVTGSWTLVGVVSWGYGCALRDFPGVYARVQSFLPWITQQMQ RFS"
3'UTR	18406. .18578 /gene="Isp2"

ORIGIN

Alignment Scores:

Pred. No.:	0.391	Length:	20247
Score:	84.00	Matches:	12
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

US-09-858-332G-15 (1-12) x AY520825 (1-20247)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
```

```
                 ||||||||||||||||||||||||||||||||||
Db       12772 CACAACCACAACCACAACCACAACCACAACCACAAC 12807

RESULT 31
YSCH9433/c
LOCUS       YSCH9433               29276 bp    DNA     linear   PLN 19-MAR-2002
DEFINITION  Saccharomyces cerevisiae chromosome VIII cosmid 9433.
ACCESSION   U11582 U00093
VERSION     U11582.1  GI:2289793
KEYWORDS    .
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 29276)
  AUTHORS   Favello,T.
  TITLE     The sequence of S. cerevisiae cosmid 9433
  JOURNAL   Unpublished (1994)
REFERENCE   2  (bases 1 to 29276)
  AUTHORS   Johnston,M., Andrews,S., Brinkman,R., Cooper,J., Ding,H., Dover,J.,
            Du,Z., Favello,A., Fulton,L., Gattung,S., Geisel,C., Kirsten,J.,
            Kucaba,T., Hillier,L., Jier,M., Johnston,L., Keppler,D.,
            Langston,Y., Latreille,P., Louis,E., Macri,C., Mardis,E.,
            Mouser,L., Nhan,M., Rifken,L., Riles,L., St.Peter,H., Thornton,L.,
            Trevaskis,E., Vaudin,M., Vaughan,K., Vignati,D., Wilcox,L.,
            Willis,A., Wilson,R., Wohldman,P. and Waterston,R.
  TITLE     The complete sequence of Saccharomyces cerevisiae chromosome VIII
  JOURNAL   Unpublished (1994)
REFERENCE   3  (bases 1 to 29276)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUN-1994) Washington University, St. Louis, MO 63110,
            USA
COMMENT     On Aug 1, 1997 this sequence version replaced gi:2289791.
            Submitted by:
                      Genome Sequencing Center
                      Department of Genetics, Washington University,
                      St. Louis, MO  63110, USA
                      e-mail: mj@sequencer.wustl.edu.
FEATURES             Location/Qualifiers
     source          1. .29276
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /strain="S288C (AB972)"
                     /db_xref="taxon:4932"
                     /map="VIII"
     gene            275. .1273
                     /gene="SNF6"
     CDS             275. .1273
                     /gene="SNF6"
                     /note="YHL025w"
                     /codon_start=1
                     /product="Probable transcription factor"
                     /protein_id="AAB65078.1"
                     /db_xref="GI:2289895"
                     /translation="MGVIKKKRSHHGKASRQQYYSGVQVGGVGSMGAINNNIPSLTSF
                     AEENNYQYGYSGSSAGMNGRSLTYAQQQLNKQRQDFERVRLRPEQLSNIIHDESDTIS
                     FRSNLLKNFISSNDAFNMLSLTTVPCDRIEKSRLFSEKTIRYLMQKQHEMKTQAAELQ
                     EKPLTPLKYTKLIAAAEDGSRSTKDMIDAVFEQDSHLRYQPDGVVVHRDDPALVGKLR
                     GDLREAPADYWTHAYRDVLAQYHEAKERIRQKEVTAGEAQDEASLQQQQQQDLQQQQQ
                     VVTTVASQSPHATATEKEPVPAVVDDPLENMFGDYSNEPFNTNFDDEFGDLDAVFF"
     gene            2073. .4214
                     /gene="YHL024w"
     CDS             2073. .4214
                     /gene="YHL024w"
                     /note="Similiar to RNA binding proteins in N terminus with
                     Asparagine rich region?"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAB65077.1"
                     /db_xref="GI:2289894"
                     /translation="MKTEISTADSLRDPPSNGLKADSELVIREDIDQFLPSEVSSLGS
                     DHQNDGEDSDTDSDNFLQDPEDDVDEESTGRGTVTTTSTSTESRGRPSSCIFVASLAA
                     ALSDDELCLSVTENFKKYGDLARVKVLRDNANRPYAFVQYNNDHDAKHALIRAQGTLL
                     NGRRLRCEPAKVNRTLYLKNQQSIDFNEISQICEKFGGLEQIVPDRTDNQYTRRYTYP
                     ISSANSWFVQFVYRDDAIRAYANLRTDPNWIIEWAQNINVPKNYNLLHKSKFKSSKYH
                     QNNGIINNDGSNNNDNNNSNNNNREDSRRNGDVIEEECGHVHGSDSEEKLTSDGIYDD
                     EDKDSEITIDKRSIFVGQLDKETTREELNRRFSTHGKIQDINLIFKPTNIFAFIKYET
                     EEAAAAALESENHAIFLNKTMHVQYKEVGGRHNRKFSGKNGGSNFNHHQFFSTRSGKT
                     FTGPELNLAPPPINMYRKMSGGSQQESETMMPYMPMGPMPMGPPPPNAASLSDFDMFP
                     PSYSTFMKGMMPLRRKSMPNSWSSPSSKSVNSENESVNGGDENSELPSEIPESSGRYN
                     AANSFTTYNNSSAGNSNNNNNNNSNSNKSQYKKRYARRSSYGYNEVPPKPYYFQPYY
                     YHPMQYHMGPMGPLHPSQGSAGNHHPYMMVYPMSPPPPSGLDGSMIPPPINVSQSHAA
                     NHGSTHVHANEFISNDTGDINEDNKAYSLDY"
     gene            complement(4547. .7987)
                     /gene="YHL023c"
     CDS             complement(4547. .7987)
                     /gene="YHL023c"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="unknown"
                     /protein_id="AAB65076.1"
                     /db_xref="GI:2289893"
                     /translation="MDECLPNSCLLGVHLVISTHSGPQIVYHYPPSNTAFLTNNPTKH
                     QHLYGNHANLNKNTSTNKEEKLFNSGSTKTASQIALNESAKSYNTAITPSMTNTNTNN
                     VTLPPTRSHANTVGSQSSIPAATNGVGYRKTDIEDTSRTFQYQETESETSSSGLSDSE
                     LSTDYLDISSDSFSISSSLSSSSLSSSPSSSSSSSSPPQDGLSRTNSSFQSTDSMSPTS
                     PQMIMENDSISVAESYLDSGTNNKSRAASKRSQNFFHKLSTKKSTDSKTHSPVRKLKS
                     KPSQSTKKGNKLLKNTSNETDGNAFTGSCSISSKKSLSSTGEHNQELRNSSLNDTPGQ
                     SPHHYHHRYHHYHKNAATSQRNSHTQYDVEEEDMEVSAMLQDGKISMNEIFFEEENFQ
                     DINKILEFDNDFVAEFCSPEREMCNTRFEFTVDNFCFLGLPIHVDSQGRWRKSKHKNK
                     TRSKRSSSTTTNISRKKSIASKISSLSENTLKKVNSGEADTVYDSNIGHEASTDTPNL
                     RINTDVSGNEFEREKEDLGKNMNMFHVCFVMNPHLIEYNKRIDDMYQFVVTRLSLLLR
                     YVQSKTSYISSECHIILKEKERVLKHSKTYQSIRGAGNKGKYLYQRILAKSSLARALT
                     ECVDKIQRNEIACLEINDDKVISLQIPIQNEFEKMPNFKLQPVLRGSYLTSILNMKFL
                     EKSSLRIESQNRQNDQAQFSDTNNNIYRFGNNINSTGHCGAANVDDGDDNESNYYCDD
                     NDDLLNYALLLLDEPNNIISSLETFSYQDDIGTIILKHLVRNIQPNIPLRSYRYLITD
                     LLDNPSSLDDLTTETNSLESSILRSCALHLMYWRHARIVIPLSSKYTYIVSPLAPIQG
                     YTIDDYKSTSQNDGNVKKMDDRENNKSGSDRVPLIYQNSMLFRSKFPSLPSLPIFLSL
                     LSTDKPQAYSNIIPSREHKPVYLNALAWLIQYGYVTQLLTFINIRVDKHIKMAVDEDL
                     EKEGFRKTNTARRPSMDYKKTDKKLDDEDGQSRDANASEACSGKNEGMQSNDNNKDVD
                     EKDNENDSRVDDRDDNEIAIADEEEILHFEYDDPEMQHDYTIILEPERATAIEKRWLY
                     RCIYGQPSDIQILFNKLLKYFNGKVPMELVIIKEEISRHDLKKLLNALDKYLIEIHHW
                     "
     tRNA            8179. .8250
                     /product="tRNA-His"
                     /note="codon recognized: CAC"
     gene            complement(8385. .9581)
                     /gene="SPO11"
     CDS             complement(8385. .9581)
                     /gene="SPO11"
                     /note="YHL022c"
                     /codon_start=1
                     /product="Sporulation protein"
                     /protein_id="AAB65075.1"
                     /db_xref="GI:2289892"
                     /translation="MALEGLRKKYKTRQELVKALTPKRRSIHLNSNGHSNGTPCSNAD
                     VLAHIKHFLSLAANSLEQHQQPISIVFQNKKKKGDTSSPDIHTTLDFPLNGPHLCTHQ
                     FKLKRCAILLNLLKVVMEKLPLGKNTTVRDIFYSNVELFQRQANVVQWLDVIRFNFKL
                     SPRKSLNIIPAQKGLVYSPFPIDIYDNILTCENEPKMQKQTIFPGKPCLIPFFQDDAV
                     IKLGTTSMCNIVIVEKEAVFTKLVNNYHKLSTNTMLITGKGFPDFLTRLFLKKLEQYC
                     SKLISDCSIFTDADPYGISIALNYTHSNERNAYICTMANYKGIRITQVLAQNNEVHNK
                     SIQLLSLNQRDYSLAKNLIASLTANSWDIATSPLKNVIIECQREIFFQKKAEMNEIDA
                     RIFEYK"
     gene            complement(9885. .11282)
                     /gene="YHL021c"
     CDS             complement(9885. .11282)
                     /gene="YHL021c"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="unknown"
                     /protein_id="AAB65074.1"
                     /db_xref="GI:2289891"
                     /translation="MLRSNLCRGSRILARLTTTPRTYTSAATAAAANRGHIIKTYFNR
                     DSTTITFSMEESSKPVSVCFNNVFLRDASHSAKLVTTGELYHNEKLTAPQDIQISEDG
                     KSLVVKWKDGGHHQFPLQFFIDYKGSSFVSPATRKQESRYRPQLWNKRILKDNVKDLL
```

```
                     SVSYNEFIDPKDDSKLFQTLVNLQKFGIAFISGTPSSSSEGLTIQKICERIGPIRSTV
                     HGEGTFDVNASQATSVNAHYANKDLPLHTDLPFLENVPGFQILQSLPATEGEDPNTRP
                     MNYFVDAFYATRNVRESDFEAYEALQIVPVNYIYENGDKRYYQSKPLIEHHDINEDNT
                     LLGNYEALIKCINYSPPYQAPFTFGIYDKPSDLNNNLDLNLITTPAKLTERFLFKSFI
                     RGLNLFESHINDFNNQFRLQLPENCCVIFNNRRILHANSLTSSNQQWLKGCYFDSDTF
                     KSKLKFLEEKFPHDK"
     gene            complement(11665. .12879)
                     /gene="OPI1"
     CDS             complement(11665. .12879)
                     /gene="OPI1"
                     /note="YHL020c"
                     /codon_start=1
                     /product="OPI1 (Negative regulator of phospholipid
                     biosynthesis)"
                     /protein_id="AAB65073.1"
                     /db_xref="GI:2289890"
                     /translation="MSENQRLGLSEEEVEAAEVLGVLKQSCRQKSQPSEDVSQADKMP
                     ASESSTTPLNILDRVSNKIISNVVTFYDEINTNKRPLKSIGRLLDDDDDEHDDYDYND
                     DEFFTNKRQKLSRAIAKGKDNLKEYKLNMSIESKKRLVTCLHLLKLANKQLSDKISCL
                     QDLVEKEQVHPLHKQDGNARTTTGAGEDETSSDEDDDDEEFFDASEQVNASEQSIVVK
                     MEVVGTVKKVYSLISKFTANSLPEPARSQVRESLLNLPTNWFDSVHSTSLPHHASFHY
                     ANCEEEQKVEQQQQQQQQQQQQQLLQQQLLQQQQQKRNKDGDDSASPSSSVTANGKVLI
                     LAKESLEMVRNVMGVVDSTLGKAEEWVKQKQEVKEMIRERFLQQQQQYRQQQQKDGNY
                     VKPSQDNVDSKD"
     gene            complement(13154. .14971)
                     /gene="YHL019c"
     CDS             complement(13154. .14971)
                     /gene="YHL019c"
                     /note="Similiar to medium chain of clathrin coat
                     associated proteins"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAB65072.1"
                     /db_xref="GI:2289889"
                     /translation="MSSSLFILDENLEPLVSKNIRALPNLSSVLSSFKQCYHDGSPPI
                     LSQNDWFFIHLKRDFLHFVSVIHTTDKPNIDLMTILAFLEQFYHLLQKYFEIEVLTKN
                     VILDNILLVLELIDECIDFGIVQVTDPSIIKDYIRVKVNVPRVTVDNEEWSPGEESSS
                     SSGSDSDSEYSNTNKRKDKKKKRKKKKGTKGKSVGKSKLKSIMVNNKENRGINVVETV
                     KETLRNKNDTGKEAANDELPNDGNDLYINGDIAKTIIMPISWRTKGIHYAKNEFFLDV
                     IERVQYLMDFEKGVIRKNLIHGEIVCRCYLSGMPKLKISINKILNRDPQFMSNSSFHQ
                     CVSLDSINTIEKDEEKNSDDDAGLQAATDAREIEFIPPDGEFVLCQYELKRHVKDAPM
                     VRLKDFEIKPKLKKFKIQIVTKIQTNFKPTNSTSKLNVRIPLTKVFQEYKIDLSKQIR
                     FKANIGKVVFNLSDDFLLWEIQTMKGHREHSTNKSSQYNSDEDDPNTCASMVAEFPLF
                     NQEEYDRLQEEMKTSMNPPPLRTGPRLEELYRQVHDQQTSHVTPRDKLVNIDFEIPYC
                     TCSGLKVEYLKVEEPQLQYQSFPWVRYKTVSDEEYAYIV"
     gene            15131. .15493
                     /gene="YHL018w"
     CDS             15131. .15493
                     /gene="YHL018w"
                     /note="Weakly similar to pterin-4-alpha-carbinolamine
                     dehydratase (human)."
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAB65071.1"
                     /db_xref="GI:2289888"
                     /translation="MHNKIVRIASSALTGGKLLEKLKPLTRWEVQWDPNKTKCLGITR
                     EVTFKDYETTWAFLTRVSMRSHLWGHHPLIHTSYTWVKLELHTHDIDPKDGAHSQLSD
                     IDVRMAKRIDSYIDEMTT"
     gene            15699. .17297

Alignment Scores:
Pred. No.:              0.557          Length:         29276
Score:                  84.00          Matches:        12
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     15             Gaps:           0

US-09-858-332G-15 (1-12) x YSCH9433 (1-29276)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      11588 CATAATCATAATCATAATCATAATCATAATCATAAT 11553
```

```
RESULT 32
AF309805
LOCUS       AF309805               32083 bp    DNA     linear   PLN 14-MAR-2001
DEFINITION  Pneumocystis carinii f. sp. carinii glutathione synthetase Gsh1
            (gsh1) gene, partial cds; putative a-factor pheromone receptor Ste3
            (ste3) gene, complete cds, alternatively spliced; and Ydr124wp-like
            protein, helix-loop-helix DNA binding protein, Yer049wp-like
            protein, protein kinase Ste20 (ste20), putative sucrose carrier
            Sca1 (sca1), putative Na+/H+ exchanger Nhx1 (nhx1), putative 26S
            protease regulatory subunit 4 (mts2), hypothetical protein, glycyl
            tRNA synthetase, and coiled-coil protein genes, complete cds.
ACCESSION   AF309805
VERSION     AF309805.1  GI:11596246
KEYWORDS    .
SOURCE      Pneumocystis carinii
  ORGANISM  Pneumocystis carinii
            Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;
            Pneumocystidaceae; Pneumocystis.
REFERENCE   1  (bases 1 to 32083)
  AUTHORS   Smulian,A.G., Sesterhenn,T., Tanaka,R. and Cushion,M.T.
  TITLE     The ste3 pheromone receptor gene of Pneumocystis carinii is
            surrounded by a cluster of signal transduction genes
  JOURNAL   Genetics 157 (3), 991-1002 (2001)
   PUBMED   11238389
REFERENCE   2  (bases 1 to 32083)
  AUTHORS   Smulian,A.G., Sesterhenn,T.M., Tanaka,R. and Cushion,M.T.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-OCT-2000) Infectious Disease Division, Department of
            Internal Medicine, University of Cincinnati College of Medicine,
            231 Bethesda Avenue, Cincinnati, OH 45267-0560, USA
FEATURES             Location/Qualifiers
     source          1. .32083
                     /organism="Pneumocystis carinii"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4754"
                     /clone="cosmid W15A6"
                     /note="forma specialis: carinii"
     gene            <1. .900
                     /gene="gsh1"
     mRNA            join(<46. .339,379. .701,753. .900)
                     /gene="gsh1"
                     /product="glutathione synthetase Gsh1"
                     /note="5' boundary not determined"
     CDS             join(<46. .339,379. .701,753. .844)
                     /gene="gsh1"
                     /note="5' boundary not determined"
                     /codon_start=2
                     /evidence=not_experimental
                     /product="glutathione synthetase Gsh1"
                     /protein_id="AAG38537.1"
                     /db_xref="GI:11596248"
                     /translation="SHGIQAYRVSLHSILDSIELRQNDKALLFKSPNSKIEEISVVYF
                     RTGYSPSHYPTSIQWSARLLIEESRAIKCPTVITQLAGCKKVQQVLCENNIINKFLPE
                     NISKSLKETFVSIYPLNKSLSGQEAYHLLMNNPQNYVLKPQREGGGNNIYGLNILSFL
                     ETISESQREQYILMEMIHPFLQTNTIVQNNELYHTNGGNELGIYGGYLLRTKDKDKNE
                     GGVMIGCSSIDSCVLIP"
     gene            1058. .2865
                     /gene="ste3"
     mRNA            join(1058. .1887,1934. .2411,2463. .2865)
                     /gene="ste3"
                     /product="putative a-factor pheromone receptor Ste3a"
     mRNA            join(1058. .1887,1934. .>2415)
                     /gene="ste3"
                     /product="putative a-factor pheromone receptor Ste3b"
     CDS             join(1359. .1887,1934. .2411,2463. .2775)
                     /gene="ste3"
                     /note="alternatively spliced"
                     /codon_start=1
                     /evidence=experimental
                     /product="putative a-factor pheromone receptor Ste3a"
                     /protein_id="AAG38536.1"
                     /db_xref="GI:11596247"
```

```
                /translation="MGEAFYIFFCLIGFLCSIIPSIWHWKYRNVAPLCLIFWVSSTNL
                VYFINSIIWYNGSETSYRGDLYCDIVTKLILGSVTGELGATAAITHYLSKIMKPSYSF
                LQSKITRRNQAIEDLLFSFGPPIMIMSLHYIVQPARYVIDGTSGCMPWTDRSWLAVAI
                VLLWPPVFGSISAYYSVKVIISYIKKRNEFQTVLKDSKTSMTLSRFIRLIGLSSLIIT
                IYLPLNIYLLIANIAEIIRSNIKYSWSHVHNWSSIIFYVSKSNMPFNRWLSPSSGIII
                FIFFGMGSDAIVMYKEIAKKLYITHFFHFIQRKIFRKKTQDIKDAENYYNSYSFEKSL
                GSCPPLFYNQTRDIRIVENGSLNECPSYPPIYTDHNNHNNNHNNHNHNNNHNHNHNHN
                HNHNHNHNHNHNNHNHNNHNHNITPSTYNQSHEHIFNTEYNKYRSHK"
CDS             join(1359. .1887,1934. .2415)
                /gene="ste3"
                /note="alternatively spliced"
                /codon_start=1
                /evidence=experimental
                /product="putative a-factor pheromone receptor Ste3b"
                /protein_id="AAG38548.1"
                /db_xref="GI:11596259"
                /translation="MGEAFYIFFCLIGFLCSIIPSIWHWKYRNVAPLCLIFWVSSTNL
                VYFINSIIWYNGSETSYRGDLYCDIVTKLILGSVTGELGATAAITHYLSKIMKPSYSF
                LQSKITRRNQAIEDLLFSFGPPIMIMSLHYIVQPARYVIDGTSGCMPWTDRSWLAVAI
                VLLWPPVFGSISAYYSVKVIISYIKKRNEFQTVLKDSKTSMTLSRFIRLIGLSSLIIT
                IYLPLNIYLLIANIAEIIRSNIKYSWSHVHNWSSIIFYVSKSNMPFNRWLSPSSGIII
                FIFFGMGSDAIVMYKEIAKKLYITHFFHFIQRKIFRKKTQDIKDAENYYNSYSFEKSL
                GR"
mRNA            complement(join(<7254. .8125,8261. .8431,8475. .>8889))
                /product="Ydr124wp-like protein"
CDS             complement(join(7254. .8125,8261. .8431,8475. .8889))
                /note="similar to Saccharomyces cerevisiae ORF Ydr124wp"
                /codon_start=1
                /evidence=not_experimental
                /product="Ydr124wp-like protein"
                /protein_id="AAG38543.1"
                /db_xref="GI:11596254"
                /translation="MSDIMKSVLGMNESLIPKEKSISLNSTWDPQNFQKKKNMNKDHL
                CIDVNTNDKSNEYDSWKTRYRSLVIGDTAEVIAFMESRFKQLQQSVCKIVAKAWIKVI
                EPKKQTRYPYNRGDRSKPSWWPENIRHKEPDHLMKSERISLLITMLRCSKIPINRLEL
                ATAEISTFIPPDKTSLLREIYHVAREEEKMRNGDIVKMGLRLLLLRQNHRSNRIKVIQ
                MKLNRLIIYQSYRNIFVLIVLQMNKIPVYTKNHFSRIHVILSKWRINYIQNQNRLSSE
                YLAYKKQKIKDTSNHTDAYNIEKDNNVNCQSLDNLSALKKNMNDSYHPMNFSDLKEDN
                QNCYTNTSMFPLSSDQFFDSPYNFSLLDKNLLMHPKHLLHNYSSSSGSTSNAAISTYL
                TTDLFSQKNHQQLSEMFSPNQFTSKYCQSYSNINHSCFPLHPDTFYNDGLPGTCFSMP
                IKTSNFPDSTGTLQIKYGQDNTFKKTDESFVEDTN"
mRNA            complement(join(<9655. .10008,10050. .10319,10361. .>10390))
                /product="helix-loop-helix DNA binding protein"
CDS             complement(join(9655. .10008,10050. .10319,10361. .10390))
                /note="similar to Schizosaccharomyces pombe protein YAWC"
                /codon_start=1
                /evidence=not_experimental
                /product="helix-loop-helix DNA binding protein"
                /protein_id="AAG38544.1"
                /db_xref="GI:11596255"
                /translation="MSLYLILIILVDKVYRTEDALSKEFHGLSEITRYTGMNGEEEEE
                EESGGIYSVESGEEILKSGVIRRDEEIHDAGGHKPIVGSEEWHRVRRSNHKVRVKREV
                ERRRRETINESISELSRIVPGCEKNKGKILERTVSYIQQLKENEANNIEKWTLEKLLT
                DQAISELSATNQKLKAECDRLWREVSVWKQAATCGINNDKDEKKDDPISNVRAEKMD"
mRNA            complement(join(<10594. .10818,10890. .11744,11787. .11975,
                12020. .12181,12434. .>12475))
                /product="Yer049wp-like protein"
CDS             complement(join(10594. .10818,10890. .11744,11787. .11975,
                12020. .12181,12434. .12475))
                /note="similar to Saccharomyces cerevisiae ORF Yer049wp"
                /codon_start=1
                /evidence=not_experimental
                /product="Yer049wp-like protein"
                /protein_id="AAG38547.1"
                /db_xref="GI:11596258"
                /translation="MFFLILRMNKVQHRSSDFITAISQEYHSARPYKHSVLTGIICEN
                LLRKIRHEIIENLAFSEKETDIYKVFQTGDLENLSGLNENERNKLCSLIELRNALYLM
                NFGYFISKVTGCGYLSGKKIDISVNMYGQGGHLLNHDDVIGSREVSFILYLIDPDETW
                LPSYGGALRLYPTVRPCFPESEHEKAIFPVWNQLCFFQVQPGRSFHDVGKKCMKRING
                GSVFQDGFHRPQPGEEGWKESEKSEEKVLSTLQNLHSNIQDYDEPRMEPRICSITFND
                EFNEFLKGHLVYLKDWINPVYLNYEKMIQLRETFLDNSVLELPDFLESTVSSRIKDWI
                IDIEKSSSTHLSLDTSSPWKLAKPLHKHRYLYIDTPDSSFNSPLISLLQNLFQSSFFS
                SMDSFLHHFNSFNFYFTYSAAFVLPPSGHWNHGQNGGYCLYMEPNTITHSTLLYSTPS
                WNNLHLILRDPGVLHFIKYVSLYSNFSRWDILSNMYFITS"
gene            complement(<13024. .>15664)
                /gene="ste20"
mRNA            complement(join(<13024. .13095,13144. .13478,13527. .13617,
                13683. .13915,13970. .14909,14960. .15039,15255. .15360,
                15410. .>15664))
                /gene="ste20"
                /product="protein kinase Ste20"
CDS             complement(join(13024. .13095,13144. .13478,13527. .13617,
                13683. .13915,13970. .14909,14960. .15039,15255. .15360,
                15410. .15664))
                /gene="ste20"
                /codon_start=1
                /product="protein kinase Ste20"
                /protein_id="AAG38545.1"
                /db_xref="GI:11596256"
                /translation="MAHNLSSQYFMNPGPAPAPPIKIPSYSSSNIYAGRCFDVSHTYK
                GSKGVLKAGWANVKEEGIKSWIWSKKYLVLTDKVLRVMKNESSTSPGLSIPLREIMNV
                SRAELKDYCFEILRDPGIGSGTKKSYFISLKSDNDLYSWMDEIYSRCPLMMVSNPTNF
                THKVHVGFDPKSGGFTGLPEEWTRLLTASAITKEDYAKNPEAVIEVLQFYSDTQLREK
                EEFGEPSTLGPDLYNIENQDIKSLEQNKMTMEKHYVNNGFNSQYPVSSTSKGAVPLRE
                WLKKMDIKDSEKMVPQRVPPALPSNLPTRLNTDMASFPHESDSFLTCGSSNVDKGFGV
                TAKKRQNVPLETLKHKQSQNNLAFISQAPTQPLPKIPTSSKNIPDIPTTFPETTMKQP
                NATKLPLSLSNARAMKQEPPSDEKVNDLTGSQVMERLRNVVSTNDPNPLYTKIKKVGQ
                GASGSVYVARVAIKQMDIAHHPRRELIVNEILVMKESHHPNIVNFLDAYLAKNSELWV
                IMEYMEGGALTDIIDNNTLTEDQIAAISLEGLQHLHPTKIFHRDIKSDNVLLDFHGRV
                KITDFGFCAKLTEQKNKRATMVGTPYWMAPEVVKQKEYGAKVDIWSLGIMAIEMIETE
                PPYLNEEPLKALYLIATNGTPTLKKPDKLSKELKSFLAVCLCVDVKSRATAEELLEHD
                FLKKSCSLSNLSSLLAFKSGK"
gene            complement(<15925. .>18086)
                /gene="sca1"
mRNA            complement(join(<15925. .16243,16294. .16567,16643. .16829,
                16868. .16923,17004. .17098,17185. .17443,17514. .17863,
                17926. .>18086))
                /gene="sca1"
                /product="putative sucrose carrier Sca1"
CDS             complement(join(15925. .16243,16294. .16567,16643. .16829,
                16868. .16923,17004. .17098,17185. .17443,17514. .17863,
                17926. .18086))
                /gene="sca1"
                /codon_start=1
                /product="putative sucrose carrier Sca1"
                /protein_id="AAG38546.1"
                /db_xref="GI:11596257"
                /translation="MIGLYAHVNASKDHFDDVINASPPLEEPQEPPLRGTLYLLFLTL
                VLGGLQLTCKSRKIFILYYIREGTPYLLSLGLTKHYTSFVWIAAPLTGIIIQPLIGFF
                SDRSRSKFGRRRPFLVVGTLCVVISLIGLAYFKDIAFFFFSGHEKVAKNMTIIVSIII
                VYLLDFSINIVQASSRALIIDMVPMVQQDLANAWASRMIGIFNVVGYLNGYLNLPKIA
                PALGNTEFKALSVIGSIVLIVCMAVTFFVAKEKRYFMESLFNFFKTIFSTIITLPSEI
                LWICIIQFFAWIELISMVFKVSDSFLCKIFFGTTYVAEIYAEKYCNKYLKYPDHLYTK
                EKEYLLIEEATRYGSFAFCAFFYYFFIRNIFLSFIFFMDFWSFYFYFMHAFNHTSKIN
                HFFFPLFVSGKKLFKYIFRLKFYNILQNINIQFLWTFGHFIFTLCMLSTILVKSITQA
                TIIIALCGLSWAITLWAPFSLIAIQLSSNNQLHRSGTILGVHNIFITVPQVLSIIMVG
                IIFKLTGYKKFEDIIKCNDNMDYSFIWIFQISGISALIAMYLSFKIYTNDPSYERNKI
                "
gene            <18200. .>20207
                /gene="nhx1"

Alignment Scores:
Pred. No.:                0.608          Length:          32083
Score:                    84.00          Matches:         12
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       15             Gaps:            0

US-09-858-332G-15 (1-12) x AF309805 (1-32083)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db       2608 CATAATCACAACCATAATCACAATCATAATCACAAC 2643

RESULT 33
```

```
AC115607/c
LOCUS       AC115607               64707 bp    DNA     linear   INV 21-FEB-2004
DEFINITION  Dictyostelium discoideum chromosome 2 map 4559693-4624405 strain
            AX4, complete sequence.
ACCESSION   AC115607 AC116959
VERSION     AC115607.2  GI:28829478
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 64707)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 64707)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 64707)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 64707)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   5  (bases 1 to 64707)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2004) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On or before Feb 21, 2004 this sequence version replaced
            gi:28850456, gi:19570077.
            CDS predictions from GeneID may contain errors. Further Information
            is available from IMB Jena, Department of Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES             Location/Qualifiers
     source          1. .64707
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /chromosome="2"
                     /map="4559693-4624405"
     CDS             1336. .4150
                     /note="GeneID exon scores (in order of location ranges):
                     199.30 - GSCJ_ID dd_01223"
                     /pseudo
                     /codon_start=1
     CDS             complement(join(4548. .4688,4778. .4889,4959. .5230,
                     5391. .5624,6033. .6154,6498. .6564))
                     /note="GeneID exon scores (in order of location ranges):
                     10.05, 6.71, 31.71, 28.25, 17.96, 5.94 - GSCJ_ID dd_01225"
                     /codon_start=1
                     /product="similar to Homo sapiens (Human). Hypothetical
                     protein FLJ11197"
                     /protein_id="AAL92367.2"
                     /db_xref="GI:28829479"
                     /translation="MGQLLSFINGNDHTEQIFIDFEHAQPSDDERELHKTVNEVLIRG
                     PAIIDKLLAYAGCNEFIRRAITNPGPETEDAAWEAVLPSVDQLQEFYDFSLELETCFP
                     KLLVALCKNDPKASLSNQQALAKQLADIFDFVLKFDDAKMVNPAIQNDFSYYRRTLNR
                     MKLTKKDANIKIRDELANRMSLFFAYPTPMMKVLSETTVKFLSQDTTVPRDNVTTALA
                     TMANVCHDMVEKKKFTQDDLNMFCLRAMVGSIILFDHIHPQGAFVKKSPVNIKPCIVT
                     LKDSNTQSSPGLLNALRFTTIHLNDVDTPGAIKQLLLLD"
     CDS             7749. .9245
                     /note="GeneID exon scores (in order of location ranges):
                     107.75 - GSCJ_ID dd_01227"
                     /codon_start=1
                     /product="similar to Homo sapiens (Human). similar to
                     RIKEN cDNA 2810423E13 gene"
                     /protein_id="AAL92368.2"
                     /db_xref="GI:28829480"
                     /translation="MSYKIINSSDDDNDNNNNNDINNNENIIREEEKNKTLPNSPIQT
                     DEQKKKQTLSINIMVLTVFLQSIGFTLILPSMLNYLNSNDPQFKRYFGYIVALYSLGQ
                     FLGSPLFGKWSNKRAASEPVVISIVISIIGSIFYSICYKFHGFAFPAIMGTARFIVGF
                     GAGNVSVCRAYATETATIENKTQIMGKMSGAQGAGFVLGPGIGFLLNFCNFTIGQMVI
                     NKYTAPGYLSILLAILNIILVLVSFRDARDLEKRNLKTPETASLLSINDADADADANN
                     DADKPKKNMEQRLKPIVLSIFLFMVVITIFAVFETVLTLMTLQYYDWGSTENYIILGT
                     SGIISVGIFIAISSPVIKKFDDRKTALFGFTCLFIALVLLINYNIPIGSRMTLPKWQF
                     FMGSAFVSVGYPIASSLVYAIFSKVLNPNSHTQGTKMGWLTAGGSLARMVGPLWSQEI
                     WTATNQQGMELFFVTAVITLIAIVVFIIFYKDLSPHPEFVSQQKTVNK"
     CDS             join(11117. .11358,11490. .12591)
                     /note="GeneID exon scores (in order of location ranges):
                     13.04, 88.43 - GSCJ_ID dd_01228"
                     /codon_start=1
                     /product="similar to Homo sapiens (Human). Testis
                     intercellular mediator Peas (Sortilin 1) (Hypothetical
                     protein) (Testis intracellular mediator protein)"
                     /protein_id="AAL92369.2"
                     /db_xref="GI:28829481"
                     /translation="MTNNSNIENCNTIYDKQITQFKINRNGSSGCGSSNESGSLLISI
                     KIEHILEFPDEVLLTILSFLEPNDIKNIYLSSRYLSRFCFENKIWKAICKRKWSSSPI
                     FKRKPITNWRSYYLKKKGFMDLSGGLSWIEIFSSSGDKPSPRYQNTCTSVGNEIYFIG
                     GQDHPENRFNTIHKFNCTTQEFTRLIPTGSSPPKFARHTSTVIGSKIYSHGGFDGFSQ
                     HFGLFVYCTIENRWETLSPTGDIPISRTNHASTSIGDNFYIFGGMYKNGSDLVFLNDL
                     FEYNTIENKWTQLKGTGDHPPQKCGHKLFNFGGKLLLFGGGYGSQWDIKYNDVHIYDR
                     SLNRWTKVKTKGEINVCTFTVSWSVGPFLFIYGGQSVKDDLLTNDLYMLDTVNMEWSS
                     INTNCTPFPRDMGSGVLVGSTLYMFGGFSSTPLDSLFSLDLLKDENLNISPSPNI"
     CDS             complement(12707. .14728)
                     /note="GeneID exon scores (in order of location ranges):
                     186.10 - GSCJ_ID dd_01229"
                     /codon_start=1
                     /product="similar to Dictyostelium discoideum (Slime
                     mold). TRFA"
                     /protein_id="AAL92370.2"
                     /db_xref="GI:28829482"
                     /translation="MKNYNNNTAHLKALQAEISKKKFIRRYLCKKIVEDRIEDQYNKK
                     EQEQKEIEQKLKLEQEQKEIELKKQQQSLLTEEQRLQQEQQEMEQRQKTITLKREKLQ
                     KKVNSLKDLQQQQQSKIIVQKEILETLESNKNTLFLLLKTVMAEERRQTEIQKALFLK
                     KQQQQQQQFLYQQQQQQQQQQQQQQQQQQQQQQQQHDQQQQQHDQQQDQQKQHDQQQHQEQ
                     QPQQQQQFNNNPNFNGNTTNNSNQFMNGQNIQFNNNDIHQSQPQPQSQPQPQPQPQPQP
                     QPQPQPQPQPQPQPQQPQPQPQPQQQQQLQFSSNNNGTFNNTNNYNNGSFNNNNNNNN
                     NNNNNNSGFINSSNGNNFNSYNNNSGNSNNNNNNNNNNVSQYSNGVNGNNNYNNGF
                     NNNINGSIDNNNNNNTTNSNNVYNYQNKSYSVDDRSSYNHKPNFMQPSSPSMQGQYYP
                     STKPITGMSPLSPFSSPVSPSQTLFINNSNSNSNVNSFNNNNNNNNSSNNSNNSNNSN
                     RISPNLTSSNMYRNDKDQFYNSNNSNKGNYNNANNNSNNNNNRFYEKDRDQYNNNNIN
                     NNNNNNNNNSNNNNNNNNNNNNNNNNNINNNNNNNGNGYHYQNYNNNDNRGYNNNRR
                     SDNDNSYEDGSSNNNNYNNNNRGHPYKNNYHHNNNRRIYNNNNWKSQTN"
     CDS             16507. .16674
                     /note="GeneID exon scores (in order of location ranges):
                     3.53 - GSCJ_ID dd_01220"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO52016.1"
                     /db_xref="GI:28829483"
                     /translation="MEKIKKQNNIGSYSSDKDGEYKKFRFSKTSIRGFLLKEYEREVP
                     IIVKIQVMYII"
     CDS             join(16869. .17233,17321. .17990)
                     /note="GeneID exon scores (in order of location ranges):
                     28.66, 68.01 - GSCJ_ID dd_01221"
                     /codon_start=1
                     /product="similar to putative sphingoid base-phosphate
                     phosphatase [Schizosaccharomyces pombe]"
                     /protein_id="AAO52017.1"
```

```
                     /db_xref="GI:28829484"
                     /translation="MASILGEEVFFILALPISTWCVATQLGVELCVVLALTIGGGNIL
                     KNTFTLPRPPPNIVWTNTAHQKDHGLPSTHTASAFGLTFYFLIYTYFLFPTIGESFNI
                     SLLSMFFIVLFWSTSVMFSRLYNGHHTPMDVIAGLIVAITSILATTYQLRYFIEGMVL
                     SETFLFGPMLYIAILSAILFFHPQANTGPTPAYPETGLVCGASLGSLISLWLHAQHPC
                     PLMNQELLLLEANYDSIVSTIHSIPLLLHGSRILIGLVLVGIAKVFSKKFFFFAYDLV
                     IRANTNNEQSQPITTVSFDPNKKIIVTPTIEAFSKLFVYTCVSFTIVSMPYLFYYLNI
                     QTSADVTRYY"
     CDS             join(18447. .18472,18540. .18883,19090. .19319,19444. .19559,
                     19677. .21864)
                     /note="GeneID exon scores (in order of location ranges):
                     -0.96, 4.97, 2.47, 5.06, 92.96 - GSCJ_ID dd_01222"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO52018.1"
                     /db_xref="GI:28829485"
                     /translation="MKIYQPFINNSITVFFQRNSKHDFLLMVYYSLLFTSSFCIGSVL
                     LVISSIRHFFLAFELFWLLVIATTAIVSSEHFYGIIKNVIRLCVAVIIGSILSYISFS
                     MVGRSVWPNCFIFFVILLLGTETFRKEFYSPIDNLLFTLGSAASFGSCLLTCLLLRAP
                     FANILFIQTSNTLLKRSISYIKQLKLQTNLVFSRNNNFKIYYRDIDENSKKKTNDLIL
                     KLQFQHQDYSERLNHLDQMLSDSKKEFWAKNYIQHFERIQILFERNLKKLIAVGLKIN
                     DEIDDKIRNQIEPIIPQVNILFDEILKIFEEMGNQLKYKFIHTNNNDNSQWFKNTIAS
                     DYYKNLDKRDENYHSFFYCHLDFVYKINGLKDIKENSLEWNQIKINKSFEIIDETISK
                     IDQLFESIFKEAPLFNPNVDNSLSKVYILLKCFNTFCLEQKLLSKEIFLVSYHNSLNR
                     NKPQEYLAIVATIKRIWKYTKTYKQRKLEKLENERNTNEDTTPIKEKFKLYFMNFLKD
                     RILYNWQFTIKYSLVSAGGAVAIFEIKKHTKFQLFEMLNWMLNAYIVTAGPDIGALGA
                     LTSVRVIVLIIGGFLGYASIIICEFGGNDSSKALIYIGFSFAIIFLLGIILNFQIFKG
                     MITNIVFTFSVVSIPLYENGSNIIFTLYRIGYVVLGFILVFILSMIIPYYDYRELEKN
                     LFKIPFLIIELVDYFFYYSFEQPIHESNKTLLSIPYNENNENYVFNNIDYNNFSHVNN
                     INSEKSFKISEMITIDQYKSVFSEKYINLRRIFPAQRNLLLDSKLELIFKKSEFKNLE
                     NDLLSIQYFHDLFGVLEFAFVESHETHIRMIGKETRIQARRLFDELELSSTLLCHIQE
                     ERDNLGRVIRSSSLPYISQKRDNGIDLMNQLIDHVNSQISNYNIPTLEKHQLNATLFA
                     LHSFVKQYHLIFKKLILKKESMVNNSKVFILNNNNYKINSNMIDVKESIFLNQ"
     CDS             complement(join(22090. .22485,22601. .22978,23168. .23311))
                     /note="GeneID exon scores (in order of location ranges):
                     42.25, 39.59, 19.88 - GSCJ_ID dd_01224"
                     /codon_start=1
                     /product="similar to Dictyostelium discoideum (Slime
                     mold). Protein phosphatase 4 catalytic subunit (EC
                     3.1.3.16) (Serine/threonine protein phosphatase)"
                     /protein_id="AAO52019.1"
                     /db_xref="GI:28829486"
                     /translation="MSSDLDRQIEQLKRCEIIKESEVRALCSKAREILLEEGNVQRVD
                     SPVTICGDIHGQFYDLKELFKVGGDCPQTNYLFMGDFVDRGFYSVETFLLLLALKVRY
                     PDRITLIRGNHESRQITQVYGFYEECVRKYGSVTVWKYCTEIFDYLSLSALVDGKIFC
                     VHGGLSPSINTLDQIRAIDRKQEVPHEGPMCDLMWSDPEDIPGWNGSPRGAGFLFGED
                     VVQKFNHDNNLEFICRAHQLVMEGFKYMFNETLVTVWSAPNYCYRCGNVAAILQLDEN
                     LKKNFAIFEAAPQESRGAPAKKPAPEYFL"
     CDS             24836. .25615
                     /note="GeneID exon scores (in order of location ranges):
                     49.65 - GSCJ_ID dd_03410"
                     /codon_start=1
                     /product="similar to Plasmodium falciparum. Hypothetical
                     protein"
                     /protein_id="AAO52020.1"

Alignment Scores:
Pred. No.:                1.2          Length:          64707
Score:                    84.00        Matches:         12
Percent Similarity:       100.00%      Conservative:    0
Best Local Similarity:    100.00%      Mismatches:      0
Query Match:              100.00%      Indels:          0
DB:                       2            Gaps:            0

US-09-858-332G-15 (1-12) x AC115607 (1-64707)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      13956 CACAACCACAACCACAATCACAACCACAACCACAAC 13921

RESULT 34
AC113185
LOCUS       AC113185               65263 bp    DNA     linear   HTG 26-FEB-2002
DEFINITION  Mus musculus clone RP24-195K9, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC113185
VERSION     AC113185.1  GI:18921363
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 65263)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-195K9
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 65263)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-FEB-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L21360
                Center clone name: 195_K_9
            --------------.
            * NOTE: This record contains 81 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      674: contig of 674 bp in length
            *      675      774: gap of 100 bp
            *      775     1480: contig of 706 bp in length
            *     1481     1580: gap of 100 bp
            *     1581     2302: contig of 722 bp in length
            *     2303     2402: gap of 100 bp
            *     2403     3131: contig of 729 bp in length
            *     3132     3231: gap of 100 bp
            *     3232     3872: contig of 641 bp in length
            *     3873     3972: gap of 100 bp
            *     3973     4671: contig of 699 bp in length
```

```
*     4672     4771: gap of 100 bp
*     4772     5477: contig of 706 bp in length
*     5478     5577: gap of 100 bp
*     5578     6284: contig of 707 bp in length
*     6285     6384: gap of 100 bp
*     6385     7076: contig of 692 bp in length
*     7077     7176: gap of 100 bp
*     7177     7885: contig of 709 bp in length
*     7886     7985: gap of 100 bp
*     7986     8672: contig of 687 bp in length
*     8673     8772: gap of 100 bp
*     8773     9474: contig of 702 bp in length
*     9475     9574: gap of 100 bp
*     9575    10310: contig of 736 bp in length
*    10311    10410: gap of 100 bp
*    10411    11129: contig of 719 bp in length
*    11130    11229: gap of 100 bp
*    11230    11939: contig of 710 bp in length
*    11940    12039: gap of 100 bp
*    12040    12747: contig of 708 bp in length
*    12748    12847: gap of 100 bp
*    12848    13556: contig of 709 bp in length
*    13557    13656: gap of 100 bp
*    13657    14365: contig of 709 bp in length
*    14366    14465: gap of 100 bp
*    14466    15179: contig of 714 bp in length
*    15180    15279: gap of 100 bp
*    15280    15988: contig of 709 bp in length
*    15989    16088: gap of 100 bp
*    16089    16763: contig of 675 bp in length
*    16764    16863: gap of 100 bp
*    16864    17576: contig of 713 bp in length
*    17577    17676: gap of 100 bp
*    17677    18391: contig of 715 bp in length
*    18392    18491: gap of 100 bp
*    18492    19051: contig of 560 bp in length
*    19052    19151: gap of 100 bp
*    19152    19864: contig of 713 bp in length
*    19865    19964: gap of 100 bp
*    19965    20675: contig of 711 bp in length
*    20676    20775: gap of 100 bp
*    20776    21488: contig of 713 bp in length
*    21489    21588: gap of 100 bp
*    21589    22307: contig of 719 bp in length
*    22308    22407: gap of 100 bp
*    22408    23129: contig of 722 bp in length
*    23130    23229: gap of 100 bp
*    23230    23941: contig of 712 bp in length
*    23942    24041: gap of 100 bp
*    24042    24729: contig of 688 bp in length
*    24730    24829: gap of 100 bp
*    24830    25548: contig of 719 bp in length
*    25549    25648: gap of 100 bp
*    25649    26388: contig of 740 bp in length
*    26389    26488: gap of 100 bp
*    26489    27215: contig of 727 bp in length
*    27216    27315: gap of 100 bp
*    27316    28022: contig of 707 bp in length
*    28023    28122: gap of 100 bp
*    28123    28836: contig of 714 bp in length
*    28837    28936: gap of 100 bp
*    28937    29646: contig of 710 bp in length
*    29647    29746: gap of 100 bp
*    29747    30455: contig of 709 bp in length
*    30456    30555: gap of 100 bp
*    30556    31269: contig of 714 bp in length
*    31270    31369: gap of 100 bp
*    31370    32086: contig of 717 bp in length
*    32087    32186: gap of 100 bp
*    32187    32905: contig of 719 bp in length
*    32906    33005: gap of 100 bp
*    33006    33722: contig of 717 bp in length
*    33723    33822: gap of 100 bp
*    33823    34502: contig of 680 bp in length
*    34503    34602: gap of 100 bp
*    34603    35297: contig of 695 bp in length
*    35298    35397: gap of 100 bp
*    35398    36118: contig of 721 bp in length
*    36119    36218: gap of 100 bp
*    36219    36930: contig of 712 bp in length
*    36931    37030: gap of 100 bp
*    37031    37730: contig of 700 bp in length
*    37731    37830: gap of 100 bp
*    37831    38533: contig of 703 bp in length
*    38534    38633: gap of 100 bp
*    38634    39346: contig of 713 bp in length
*    39347    39446: gap of 100 bp
*    39447    40156: contig of 710 bp in length
*    40157    40256: gap of 100 bp
*    40257    40954: contig of 698 bp in length
*    40955    41054: gap of 100 bp
*    41055    41765: contig of 711 bp in length
*    41766    41865: gap of 100 bp
*    41866    42599: contig of 734 bp in length
*    42600    42699: gap of 100 bp
*    42700    43424: contig of 725 bp in length
*    43425    43524: gap of 100 bp
*    43525    44232: contig of 708 bp in length
*    44233    44332: gap of 100 bp
*    44333    45042: contig of 710 bp in length
*    45043    45142: gap of 100 bp
*    45143    45846: contig of 704 bp in length
*    45847    45946: gap of 100 bp
*    45947    46654: contig of 708 bp in length
*    46655    46754: gap of 100 bp
*    46755    47467: contig of 713 bp in length
*    47468    47567: gap of 100 bp
*    47568    48283: contig of 716 bp in length
*    48284    48383: gap of 100 bp
*    48384    49091: contig of 708 bp in length
*    49092    49191: gap of 100 bp
*    49192    49895: contig of 704 bp in length
*    49896    49995: gap of 100 bp
*    49996    50672: contig of 677 bp in length
*    50673    50772: gap of 100 bp
*    50773    51486: contig of 714 bp in length
*    51487    51586: gap of 100 bp
*    51587    52304: contig of 718 bp in length
*    52305    52404: gap of 100 bp
*    52405    53131: contig of 727 bp in length
*    53132    53231: gap of 100 bp
*    53232    53919: contig of 688 bp in length
*    53920    54019: gap of 100 bp
*    54020    54727: contig of 708 bp in length
*    54728    54827: gap of 100 bp
*    54828    55545: contig of 718 bp in length

Alignment Scores:
Pred. No.:                1.21          Length:          65263
Score:                    84.00         Matches:         12
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       14            Gaps:            0

US-09-858-332G-15 (1-12) x AC113185 (1-65263)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db       7528 CACAACCACAACCACAACCACAACCACAACCACAAC 7563

RESULT 35
AC149204
LOCUS       AC149204               75741 bp    DNA     linear   PLN 22-MAY-2004
DEFINITION  Medicago truncatula chromosome 7 BAC clone mth2-77n20, complete
            sequence.
```

```
ACCESSION   AC149204
VERSION     AC149204.1  GI:47564143
KEYWORDS    HTG.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 75741)
  AUTHORS   Town,C.D., Tallon,L.J., Arbogast,T., Althoff,R., Hine,E.,
            Monaghan,E., Smith,S.A., Utterback,T., Feldblyum,T., Koo,H. and
            Cheung,F.
  TITLE     Medicago truncatula BAC genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 75741)
  AUTHORS   Town,C.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAY-2004) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .75741
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3880"
                     /chromosome="7"
                     /clone="mth2-77n20"
ORIGIN

Alignment Scores:
Pred. No.:                1.39          Length:          75741
Score:                    84.00         Matches:         12
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       15            Gaps:            0

US-09-858-332G-15 (1-12) x AC149204 (1-75741)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db       8080 CATAATCATAATCATAATCATAATCATAATCATAAT 8115

RESULT 36
AC139431/c
LOCUS       AC139431               79776 bp    DNA     linear   HTG 03-FEB-2003
DEFINITION  Homo sapiens chromosome 17 clone RP13-680F15 map 17, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC139431
VERSION     AC139431.1  GI:28195962
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 79776)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP13-680F15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 79776)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-FEB-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L29419
                Center clone name: 680_F_15
            --------------.
            * NOTE: This record contains 89 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      806: contig of 806 bp in length
            *      807      906: gap of 100 bp
            *      907     1707: contig of 801 bp in length
            *     1708     1807: gap of 100 bp
            *     1808     2585: contig of 778 bp in length
            *     2586     2685: gap of 100 bp
            *     2686     3474: contig of 789 bp in length
            *     3475     3574: gap of 100 bp
            *     3575     4376: contig of 802 bp in length
            *     4377     4476: gap of 100 bp
            *     4477     5288: contig of 812 bp in length
            *     5289     5388: gap of 100 bp
            *     5389     6176: contig of 788 bp in length
            *     6177     6276: gap of 100 bp
            *     6277     7069: contig of 793 bp in length
            *     7070     7169: gap of 100 bp
            *     7170     7962: contig of 793 bp in length
            *     7963     8062: gap of 100 bp
            *     8063     8850: contig of 788 bp in length
            *     8851     8950: gap of 100 bp
            *     8951     9748: contig of 798 bp in length
            *     9749     9848: gap of 100 bp
            *     9849    10635: contig of 787 bp in length
            *    10636    10735: gap of 100 bp
            *    10736    11533: contig of 798 bp in length
            *    11534    11633: gap of 100 bp
            *    11634    12431: contig of 798 bp in length
            *    12432    12531: gap of 100 bp
            *    12532    13307: contig of 776 bp in length
            *    13308    13407: gap of 100 bp
            *    13408    14170: contig of 763 bp in length
            *    14171    14270: gap of 100 bp
            *    14271    15067: contig of 797 bp in length
            *    15068    15167: gap of 100 bp
            *    15168    15917: contig of 750 bp in length
            *    15918    16017: gap of 100 bp
            *    16018    16837: contig of 820 bp in length
            *    16838    16937: gap of 100 bp
            *    16938    17739: contig of 802 bp in length
```

```
*    17740    17839: gap of 100 bp
*    17840    18651: contig of 812 bp in length
*    18652    18751: gap of 100 bp
*    18752    19541: contig of 790 bp in length
*    19542    19641: gap of 100 bp
*    19642    20440: contig of 799 bp in length
*    20441    20540: gap of 100 bp
*    20541    21351: contig of 811 bp in length
*    21352    21451: gap of 100 bp
*    21452    22245: contig of 794 bp in length
*    22246    22345: gap of 100 bp
*    22346    23146: contig of 801 bp in length
*    23147    23246: gap of 100 bp
*    23247    24029: contig of 783 bp in length
*    24030    24129: gap of 100 bp
*    24130    24881: contig of 752 bp in length
*    24882    24981: gap of 100 bp
*    24982    25777: contig of 796 bp in length
*    25778    25877: gap of 100 bp
*    25878    26670: contig of 793 bp in length
*    26671    26770: gap of 100 bp
*    26771    27591: contig of 821 bp in length
*    27592    27691: gap of 100 bp
*    27692    28497: contig of 806 bp in length
*    28498    28597: gap of 100 bp
*    28598    29408: contig of 811 bp in length
*    29409    29508: gap of 100 bp
*    29509    30291: contig of 783 bp in length
*    30292    30391: gap of 100 bp
*    30392    31198: contig of 807 bp in length
*    31199    31298: gap of 100 bp
*    31299    32094: contig of 796 bp in length
*    32095    32194: gap of 100 bp
*    32195    32995: contig of 801 bp in length
*    32996    33095: gap of 100 bp
*    33096    33898: contig of 803 bp in length
*    33899    33998: gap of 100 bp
*    33999    34797: contig of 799 bp in length
*    34798    34897: gap of 100 bp
*    34898    35672: contig of 775 bp in length
*    35673    35772: gap of 100 bp
*    35773    36581: contig of 809 bp in length
*    36582    36681: gap of 100 bp
*    36682    37495: contig of 814 bp in length
*    37496    37595: gap of 100 bp
*    37596    38431: contig of 836 bp in length
*    38432    38531: gap of 100 bp
*    38532    39344: contig of 813 bp in length
*    39345    39444: gap of 100 bp
*    39445    40262: contig of 818 bp in length
*    40263    40362: gap of 100 bp
*    40363    41168: contig of 806 bp in length
*    41169    41268: gap of 100 bp
*    41269    42068: contig of 800 bp in length
*    42069    42168: gap of 100 bp
*    42169    42960: contig of 792 bp in length
*    42961    43060: gap of 100 bp
*    43061    43853: contig of 793 bp in length
*    43854    43953: gap of 100 bp
*    43954    44719: contig of 766 bp in length
*    44720    44819: gap of 100 bp
*    44820    45625: contig of 806 bp in length
*    45626    45725: gap of 100 bp
*    45726    46549: contig of 824 bp in length
*    46550    46649: gap of 100 bp
*    46650    47449: contig of 800 bp in length
*    47450    47549: gap of 100 bp
*    47550    48372: contig of 823 bp in length
*    48373    48472: gap of 100 bp
*    48473    49285: contig of 813 bp in length
*    49286    49385: gap of 100 bp
*    49386    50171: contig of 786 bp in length
*    50172    50271: gap of 100 bp
*    50272    51062: contig of 791 bp in length
*    51063    51162: gap of 100 bp
*    51163    51969: contig of 807 bp in length
*    51970    52069: gap of 100 bp
*    52070    52867: contig of 798 bp in length
*    52868    52967: gap of 100 bp
*    52968    53750: contig of 783 bp in length
*    53751    53850: gap of 100 bp
*    53851    54641: contig of 791 bp in length
*    54642    54741: gap of 100 bp
*    54742    55550: contig of 809 bp in length
*    55551    55650: gap of 100 bp
*    55651    56472: contig of 822 bp in length
*    56473    56572: gap of 100 bp
*    56573    57375: contig of 803 bp in length
*    57376    57475: gap of 100 bp
*    57476    58297: contig of 822 bp in length
*    58298    58397: gap of 100 bp
*    58398    59210: contig of 813 bp in length
*    59211    59310: gap of 100 bp
*    59311    60115: contig of 805 bp in length
*    60116    60215: gap of 100 bp
*    60216    61021: contig of 806 bp in length
*    61022    61121: gap of 100 bp
*    61122    61907: contig of 786 bp in length
*    61908    62007: gap of 100 bp

Alignment Scores:
Pred. No.:                1.46          Length:          79776
Score:                    84.00         Matches:         12
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       14            Gaps:            0

US-09-858-332G-15 (1-12) x AC139431 (1-79776)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      63185 CATAATCATAATCATAATCATAATCATAATCATAAT 63150

RESULT 37
ATT8O5
LOCUS       ATT8O5                 83029 bp    DNA     linear   PLN 16-APR-2005
DEFINITION  Arabidopsis thaliana DNA chromosome 4, BAC clone T8O5 (ESSAII
            project).
ACCESSION   AL021890
VERSION     AL021890.1  GI:2894557
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Bevan,M., Wedler,H., Wambutt,R., Bancroft,I., Mewes,H.W., Mayer,K.
            and Schueller,C.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 83029)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-FEB-1998) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
FEATURES             Location/Qualifiers
     source          1. .83029
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /chromosome="4"
                     /ecotype="Columbia"
```

```
gene            1. .1440
                /gene="T8O5.10"
CDS             join(1. .79,179. .459,515. .805,893. .1098,1176. .1440)
                /gene="T8O5.10"
                /note="Protein sequence is in conflict with the conceptual
                translation; similarity to Hypothetical protein,
                Saccharomyces cerevisiae, MNOS:S71693"
                /codon_start=1
                /product="putative protein"
                /protein_id="CAA17147.1"
                /db_xref="GI:2894558"
                /db_xref="GOA:O49703"
                /db_xref="InterPro:IPR004130"
                /db_xref="UniProt/TrEMBL:O49703"
                /translation="KLRVSAASSSSNFKKKPIIIIVVGMAGSGKTSFLHRLVCHTFDS
                KSHGYVVNLDPAVMSLPFGANIDIRDTVKYKEVMKQYNLGPNGGILTSLNLFATKFDE
                VNIPLELVQMITDFFSHCVVSVIEKRADQLDYVLVDTPGQIEIFTWSASGAIITEAFA
                STFPTVVTYVVDTPRSSSPITFMSNMLYACSILYKTRLPLVLAFNKTDVADHKFALEW
                MEDFEVFQAAIQSDNSYTATLANSLSLSLYEFYRNIRSVGVSAISGAGMDGFFKAIEA
                SAEEYMETYKADLDMRKADKERLEEERKKHEMEKLRKDMESSQGGTVVLNTGLKDRDA
                TEKMMLEEDDEDFQVEDEEDSDDAIDEDDEDDETKHYYL"
exon            1. .79
                /gene="T8O5.10"
                /number=1
intron          80. .178
                /gene="T8O5.10"
                /number=1
exon            179. .459
                /gene="T8O5.10"
                /number=2
intron          460. .514
                /gene="T8O5.10"
                /number=2
exon            515. .805
                /gene="T8O5.10"
                /number=3
intron          806. .892
                /gene="T8O5.10"
                /number=3
exon            893. .1098
                /gene="T8O5.10"
                /number=4
intron          1099. .1175
                /gene="T8O5.10"
                /number=4
exon            1176. .1440
                /gene="T8O5.10"
                /number=5
gene            1809. .4079
                /gene="T8O5.20"
gene            complement(join(1809. .1991,2084. .2137,2228. .2311,
                2409. .2623,3028. .3127,4038. .4079))
                /gene="T8O5.20"
CDS             complement(join(1809. .1991,2084. .2137,2228. .2311,
                2409. .2623,3028. .3127,4038. .4079))
                /gene="T8O5.20"
                /note="similarity to NADH dehydrogenase chain 4,
                Acanthamoeba castellanii mitochondrion, PIR2:S53834
                contains EST gb:N38447, T21918"
                /codon_start=1
                /product="putative protein"
                /protein_id="CAA17148.1"
                /db_xref="GI:2894559"
                /db_xref="InterPro:IPR007599"
                /db_xref="UniProt/TrEMBL:O49704"
                /translation="MTPDYVAAPVKRTAIISPYNLYLNPTLVVKQYQFWRLVTNFLYF
                RKMDLDFLFHMFFLARYCKLLEENSFRGKTTDFLYMLLFGATVLTGIVLIGGMIPYLS
                VSFSKIIFLSNSLTFMMVYVWSKQNPYIHMSFLGLFTFTAAYLPWVLLGFSILVGASA
                WGDFLGMIAGHAYYFLAFVYPRMTDRRPLKTPSFLKALFADEPVVIARPEDVRFAHAP
                FDEIHQD"
exon            complement(1809. .1991)
                /gene="T8O5.20"
                /number=1
intron          complement(1992. .2083)
                /gene="T8O5.20"
                /number=1
exon            complement(2084. .2137)
                /gene="T8O5.20"
                /number=2
intron          complement(2138. .2227)
                /gene="T8O5.20"
                /number=2
exon            complement(2228. .2311)
                /gene="T8O5.20"
                /number=3
intron          complement(2312. .2408)
                /gene="T8O5.20"
                /number=3
exon            complement(2409. .2623)
                /gene="T8O5.20"
                /number=4
intron          complement(2624. .3027)
                /gene="T8O5.20"
                /number=4
exon            complement(3028. .3127)
                /gene="T8O5.20"
                /number=5
intron          complement(3128. .4037)
                /gene="T8O5.20"
                /number=5
gene            4022. .10335
                /gene="T8O5.30"
CDS             join(4022. .4729,4822. .5290,5735. .5961,6060. .6125,
                6311. .6364,6732. .6848,7056. .7154,7211. .7391,7599. .7736,
                7938. .8132,8214. .8494,9384. .9606,9673. .9937,9981. .10136,
                10248. .10335)
                /gene="T8O5.30"
                /note="similarity to microtubule associated protein,
                Drosophila melanogaster, PATCHX:G1930122"
                /codon_start=1
                /product="putative protein"
                /protein_id="CAA17149.1"
                /db_xref="GI:2894560"
                /db_xref="InterPro:IPR000048"
                /db_xref="InterPro:IPR000225"
                /db_xref="UniProt/TrEMBL:O49705"
                /translation="MDENEPPCASPAPPRNPASSLLSDISNFKTPRRTSVVNSNISKS
                PYPHFFTASKQTPKSSSSNFRRPSMVHSYASRSKVSTSSRRLKAFELQQSQSSRKAEL
                TKEKNLRSLAKSLTVWLNFLPENPENCGCDPFENESGVGNLGKGKRDSGEALGNSKSV
                GVDTMWRSPKRLRNLGWCGEKKRSEIDSSLTGSKYSTLRESLRDVCSLDDLKQRMQFH
                LSLGSCKEIFDVMTRVSKNIDEGRIKMKPQCPLVTDFGMKEKAIKALMSYNQVWLRLG
                LYIIFGGDSFLSDSEVNSDQEMAFLKMVISKQFFSHDGLARSYVYNKMVEGLYRPGYY
                EALGNVILKRILLLVLVIDRAKSQSCLSLKYGIDGIDGGSPLMFSEKSSIKSSHQLIC
                GSEFDLMRVLIKQKMVVPSDNRKKNLANCRIALQYLKDAGVSLKDDEGMMITGEDVAD
                GDRELTISLLWNIFVHLQLPLLINGRLLTEEIYKVQGVEQNNQITMSTPLEMLLNWIQ
                DPGGQEGPQSVMSNTDYHDAVQNFILSQKLTALLGSFPEIGDLLEHNAVVSNQSVIIL
                LAFLSSKLIVKENLLDFTLAGYLTNSVFSQEQLNFHKLLCSSCQVQEKRYSRIRISCS
                SSEAVTIEEPDRENGEDATKRFQAIKAWWQDMANQNQISVGKANSHTLQGSLSRKCTT
                DSQRAYLLLAEIAAVIIQSNIRGLHARRKFRNKMRAICFLQAAVRTWLSVKNIQVVEK
                FNVEEVTLHLSERSANLKPVARYVKFIVDRSRFIKLRKSVSVIQKAVRRHQSNLHHEL
                KAALRIQLAWRSYKEKVISSITIQSYVRGWITRRMNRTYKFSSILIQFQVSFYPYISQ
                SHWRGYLTRKASKAQVLDLRVRMQTSAANIDDKKRLINKLLSALSELLSMKNVHNILH
                ICETLESATKYSDKCCEELVAAGAIEKLLTLIRSASRSIPDQQVSKHALSTLGHLARY
                PQMADELINTKGSIQTIFWELLRSSFSSQVFFLSFKNKEEAYFIASDVLKKICSSHKG
                VEAVRKLPALVKRLHVLVEELTRKANIEKRNVKGQTGKEKSERRLKEAIELVKLITRR
                "
exon            4022. .4729
                /gene="T8O5.30"
                /number=1
exon            complement(4038. .4079)
                /gene="T8O5.20"
                /number=6
intron          4730. .4821
                /gene="T8O5.30"
                /number=1
exon            4822. .5290
```

```
                     /gene="T8O5.30"
                     /number=2
     intron          5291. .5734
                     /gene="T8O5.30"
                     /number=2
     exon            5735. .5961
                     /gene="T8O5.30"
                     /number=3
     intron          5962. .6059
                     /gene="T8O5.30"
                     /number=3
     exon            6060. .6125
                     /gene="T8O5.30"
                     /number=4
     intron          6126. .6310
                     /gene="T8O5.30"
                     /number=4
     exon            6311. .6364
                     /gene="T8O5.30"
                     /number=5
     intron          6365. .6731
                     /gene="T8O5.30"
                     /number=5
     exon            6732. .6848

Alignment Scores:
Pred. No.:                1.52          Length:         83029
Score:                    84.00         Matches:        12
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              100.00%       Indels:         0
DB:                       15            Gaps:           0

US-09-858-332G-15 (1-12) x ATT8O5 (1-83029)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db       3476 CACAATCACAATCACAATCACAATCACAATCACAAT 3511

RESULT 38
CNS07TIW/c
LOCUS       CNS07TIW              101256 bp    DNA     linear   PLN 24-OCT-2001
DEFINITION  DNA centromeric region sequence from BAC DP35C12 of chromosome 5 of
            Podospora anserina.
ACCESSION   AL627322
VERSION     AL627322.1  GI:16444879
KEYWORDS    .
SOURCE      Podospora anserina
  ORGANISM  Podospora anserina
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetidae; Sordariales; Lasiosphaeriaceae; Podospora.
REFERENCE   1  (bases 1 to 101256)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT-2001) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     * Christian Barreau and Beatrice Turcq, Institut de Biochimie et de
            Genetique Cellulaire, UPR CNRS 9026, 1 rue Camille Saint Saens,
            F-33077 Bordeaux, France;
            * Alain Billault, Molecular Engines Laboratories, 20 rue Bouvier,
            F-75011 Paris, France;
            * Laurence Cattolico and Simone Duprat, Genoscope, Centre National
            de Sequencage, 2 rue Gaston Cremieux, CP5706, F-91057 Evry CEDEX,
            France; * Robert Debuchy, Sebastien Kicka and Philippe Silar,
            Institut de Genetique et Microbiologie, UMR CNRS 8621, batiment
            400, Universite Paris-Sud, F-91405 Orsay CEDEX, France;
            * Annie Sainsard-Chanet and Carole H. Sellem, Centre de Genetique
            Moleculaire, UPR CNRS 9061, batiment 24, allee de la terrasse,
            F-91190 Gif sur Yvette, France;
            base calling/assembly program : Phred/Phrap;
            consensus quality: 98,93% of bases at least Q40.
FEATURES              Location/Qualifiers
     source           1. .101256
                      /organism="Podospora anserina"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:5145"
                      /chromosome="5"
                      /clone="DP35C12"
ORIGIN

Alignment Scores:
Pred. No.:                1.84          Length:         101256
Score:                    84.00         Matches:        12
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              100.00%       Indels:         0
DB:                       15            Gaps:           0

US-09-858-332G-15 (1-12) x CNS07TIW (1-101256)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      49868 CACAACCACAACCACAACCACAATCACAATCACAAC 49833

RESULT 39
AP006378
LOCUS       AP006378              101752 bp    DNA     linear   PLN 22-JUL-2003
DEFINITION  Lotus corniculatus var. japonicus genomic DNA, chromosome 1,
            clone:LjT44A10, TM0223, complete sequence.
ACCESSION   AP006378
VERSION     AP006378.1  GI:31581009
KEYWORDS    HTG.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1
  AUTHORS   Asamizu,E., Kato,T., Sato,S., Nakamura,Y., Kaneko,T. and Tabata,S.
  TITLE     Structural Analysis of a Lotus japonicus Genome. IV. Sequence
            Features and Mapping of seventy-three TAC clones which cover the
            7.5 Mb Regions of the Genome
  JOURNAL   DNA Res. (2003) In press
REFERENCE   2  (bases 1 to 101752)
  AUTHORS   Sato,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAY-2003) Shusei Sato, Kazusa DNA Research Institute,
            Department of Plant Gene Research; 2-6-7 Kazusa-kamatari, Kisarazu,
            Chiba 292-0818, Japan (E-mail:ssato@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/, Tel:81-438-52-3935(ex.2337),
            Fax:81-438-52-3934)
FEATURES              Location/Qualifiers
     source           1. .101752
                      /organism="Lotus corniculatus var. japonicus"
                      /mol_type="genomic DNA"
                      /variety="japonicus"
                      /db_xref="taxon:34305"
                      /chromosome="1"
                      /clone="LjT44A10"
                      /clone_lib="LjT library"
                      /note="TAC clone:TM0223
                      synonym: Lotus japonicus"
ORIGIN

Alignment Scores:
Pred. No.:                1.85          Length:         101752
Score:                    84.00         Matches:        12
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              100.00%       Indels:         0
DB:                       15            Gaps:           0

US-09-858-332G-15 (1-12) x AP006378 (1-101752)
```

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      33052 CACAATCACAATCACAATCACAATCACAATCACAAT 33087

RESULT 40
AC109729_3/c
WPCOMMENT
  Sequence split into 5 fragments  LOCUS AC109729 Accession AC109729
     Fragment Name      Begin     End
     AC109729_0             1    110000
     AC109729_1        100001    210000
     AC109729_2        200001    310000
     AC109729_3        300001    410000
     AC109729_4        400001    427537
Continuation (4 of 5) of AC109729 from base 300001 (AC109729 Rattus norvegicus clone CH2

Alignment Scores:
Pred. No.:              1.99          Length:        110000
Score:                  84.00         Matches:       12
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     14            Gaps:          0

US-09-858-332G-15 (1-12) x AC109729_3 (1-110000)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      60238 CACAACCACAACCACAATCACAACCACAACCACAAC 60203

RESULT 41
AC120234_1/c
WPCOMMENT
  Sequence split into 5 fragments  LOCUS AC120234 Accession AC120234
     Fragment Name      Begin     End
     AC120234_0             1    110000
     AC120234_1        100001    210000
     AC120234_2        200001    310000
     AC120234_3        300001    410000
     AC120234_4        400001    444501
Continuation (2 of 5) of AC120234 from base 100001 (AC120234 Rattus norvegicus clone CH2

Alignment Scores:
Pred. No.:              1.99          Length:        110000
Score:                  84.00         Matches:       12
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     14            Gaps:          0

US-09-858-332G-15 (1-12) x AC120234_1 (1-110000)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      91217 CACAACCACAACCACAACCACAACCACAACCACAAC 91182

RESULT 42
AC141230_3/c
WPCOMMENT
  Sequence split into 7 fragments  LOCUS AC141230 Accession AC141230
     Fragment Name      Begin     End
     AC141230_0             1    110000
     AC141230_1        100001    210000
     AC141230_2        200001    310000
     AC141230_3        300001    410000
     AC141230_4        400001    510000
     AC141230_5        500001    610000
     AC141230_6        600001    630326
Continuation (4 of 7) of AC141230 from base 300001 (AC141230 Homo sapiens chromosome 16

Alignment Scores:
Pred. No.:              1.99          Length:        110000
Score:                  84.00         Matches:       12
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     14            Gaps:          0

US-09-858-332G-15 (1-12) x AC141230_3 (1-110000)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      99966 CATAATCATAATCATAATCATAATCATAATCATAAT 99931

RESULT 43
AP008214_000/c
WPCOMMENT
  Sequence split into 285 fragments  LOCUS AP008214 Accession AP008214
     Fragment Name      Begin     End
     AP008214_000           1    110000
     AP008214_001      100001    210000
     AP008214_002      200001    310000
     AP008214_003      300001    410000
     AP008214_004      400001    510000
     AP008214_005      500001    610000
     AP008214_006      600001    710000
     AP008214_007      700001    810000
     AP008214_008      800001    910000
     AP008214_009      900001   1010000
     AP008214_010     1000001   1110000
     AP008214_011     1100001   1210000
     AP008214_012     1200001   1310000
     AP008214_013     1300001   1410000
     AP008214_014     1400001   1510000
     AP008214_015     1500001   1610000
     AP008214_016     1600001   1710000
     AP008214_017     1700001   1810000
     AP008214_018     1800001   1910000
     AP008214_019     1900001   2010000
     AP008214_020     2000001   2110000
     AP008214_021     2100001   2210000
     AP008214_022     2200001   2310000
     AP008214_023     2300001   2410000
     AP008214_024     2400001   2510000
     AP008214_025     2500001   2610000
     AP008214_026     2600001   2710000
     AP008214_027     2700001   2810000
     AP008214_028     2800001   2910000
     AP008214_029     2900001   3010000
     AP008214_030     3000001   3110000
     AP008214_031     3100001   3210000
     AP008214_032     3200001   3310000
     AP008214_033     3300001   3410000
     AP008214_034     3400001   3510000
     AP008214_035     3500001   3610000
     AP008214_036     3600001   3710000
     AP008214_037     3700001   3810000
     AP008214_038     3800001   3910000
     AP008214_039     3900001   4010000
     AP008214_040     4000001   4110000
     AP008214_041     4100001   4210000
     AP008214_042     4200001   4310000
     AP008214_043     4300001   4410000
     AP008214_044     4400001   4510000
     AP008214_045     4500001   4610000
     AP008214_046     4600001   4710000
     AP008214_047     4700001   4810000
     AP008214_048     4800001   4910000
     AP008214_049     4900001   5010000
     AP008214_050     5000001   5110000
     AP008214_051     5100001   5210000
     AP008214_052     5200001   5310000
     AP008214_053     5300001   5410000
     AP008214_054     5400001   5510000
     AP008214_055     5500001   5610000
     AP008214_056     5600001   5710000
```

AP008214_057	5700001	5810000
AP008214_058	5800001	5910000
AP008214_059	5900001	6010000
AP008214_060	6000001	6110000
AP008214_061	6100001	6210000
AP008214_062	6200001	6310000
AP008214_063	6300001	6410000
AP008214_064	6400001	6510000
AP008214_065	6500001	6610000
AP008214_066	6600001	6710000
AP008214_067	6700001	6810000
AP008214_068	6800001	6910000
AP008214_069	6900001	7010000
AP008214_070	7000001	7110000
AP008214_071	7100001	7210000
AP008214_072	7200001	7310000
AP008214_073	7300001	7410000
AP008214_074	7400001	7510000
AP008214_075	7500001	7610000
AP008214_076	7600001	7710000
AP008214_077	7700001	7810000
AP008214_078	7800001	7910000
AP008214_079	7900001	8010000
AP008214_080	8000001	8110000
AP008214_081	8100001	8210000
AP008214_082	8200001	8310000
AP008214_083	8300001	8410000
AP008214_084	8400001	8510000
AP008214_085	8500001	8610000
AP008214_086	8600001	8710000
AP008214_087	8700001	8810000
AP008214_088	8800001	8910000
AP008214_089	8900001	9010000
AP008214_090	9000001	9110000
AP008214_091	9100001	9210000
AP008214_092	9200001	9310000
AP008214_093	9300001	9410000
AP008214_094	9400001	9510000
AP008214_095	9500001	9610000
AP008214_096	9600001	9710000
AP008214_097	9700001	9810000
AP008214_098	9800001	9910000
AP008214_099	9900001	10010000
AP008214_100	10000001	10110000
AP008214_101	10100001	10210000
AP008214_102	10200001	10310000
AP008214_103	10300001	10410000
AP008214_104	10400001	10510000
AP008214_105	10500001	10610000
AP008214_106	10600001	10710000
AP008214_107	10700001	10810000
AP008214_108	10800001	10910000
AP008214_109	10900001	11010000
AP008214_110	11000001	11110000
AP008214_111	11100001	11210000
AP008214_112	11200001	11310000
AP008214_113	11300001	11410000
AP008214_114	11400001	11510000
AP008214_115	11500001	11610000
AP008214_116	11600001	11710000
AP008214_117	11700001	11810000
AP008214_118	11800001	11910000
AP008214_119	11900001	12010000
AP008214_120	12000001	12110000
AP008214_121	12100001	12210000
AP008214_122	12200001	12310000
AP008214_123	12300001	12410000
AP008214_124	12400001	12510000
AP008214_125	12500001	12610000
AP008214_126	12600001	12710000
AP008214_127	12700001	12810000
AP008214_128	12800001	12910000
AP008214_129	12900001	13010000
AP008214_130	13000001	13110000
AP008214_131	13100001	13210000
AP008214_132	13200001	13310000
AP008214_133	13300001	13410000
AP008214_134	13400001	13510000
AP008214_135	13500001	13610000
AP008214_136	13600001	13710000
AP008214_137	13700001	13810000
AP008214_138	13800001	13910000
AP008214_139	13900001	14010000
AP008214_140	14000001	14110000
AP008214_141	14100001	14210000
AP008214_142	14200001	14310000
AP008214_143	14300001	14410000
AP008214_144	14400001	14510000
AP008214_145	14500001	14610000
AP008214_146	14600001	14710000
AP008214_147	14700001	14810000
AP008214_148	14800001	14910000
AP008214_149	14900001	15010000
AP008214_150	15000001	15110000
AP008214_151	15100001	15210000
AP008214_152	15200001	15310000
AP008214_153	15300001	15410000
AP008214_154	15400001	15510000
AP008214_155	15500001	15610000
AP008214_156	15600001	15710000
AP008214_157	15700001	15810000
AP008214_158	15800001	15910000
AP008214_159	15900001	16010000
AP008214_160	16000001	16110000
AP008214_161	16100001	16210000
AP008214_162	16200001	16310000
AP008214_163	16300001	16410000
AP008214_164	16400001	16510000
AP008214_165	16500001	16610000
AP008214_166	16600001	16710000
AP008214_167	16700001	16810000
AP008214_168	16800001	16910000
AP008214_169	16900001	17010000
AP008214_170	17000001	17110000
AP008214_171	17100001	17210000
AP008214_172	17200001	17310000
AP008214_173	17300001	17410000
AP008214_174	17400001	17510000
AP008214_175	17500001	17610000
AP008214_176	17600001	17710000
AP008214_177	17700001	17810000
AP008214_178	17800001	17910000
AP008214_179	17900001	18010000
AP008214_180	18000001	18110000
AP008214_181	18100001	18210000
AP008214_182	18200001	18310000
AP008214_183	18300001	18410000
AP008214_184	18400001	18510000
AP008214_185	18500001	18610000
AP008214_186	18600001	18710000
AP008214_187	18700001	18810000
AP008214_188	18800001	18910000
AP008214_189	18900001	19010000
AP008214_190	19000001	19110000
AP008214_191	19100001	19210000
AP008214_192	19200001	19310000
AP008214_193	19300001	19410000
AP008214_194	19400001	19510000
AP008214_195	19500001	19610000
AP008214_196	19600001	19710000

Alignment Scores:

Pred. No.:	1.99	Length:	110000
Score:	84.00	Matches:	12
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0

```
Query Match:            100.00%       Indels:        0
DB:                     15            Gaps:          0

US-09-858-332G-15 (1-12) x AP008214_000 (1-110000)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      60490 CACAACCACAACCACAACCACAATCACAACCACAAT 60455

RESULT 44
CR382132_16/c
WPCOMMENT
  Sequence split into 40 fragments  LOCUS CR382132 Accession CR382132
     Fragment Name      Begin     End
     CR382132_00            1    110000
     CR382132_01       100001    210000
     CR382132_02       200001    310000
     CR382132_03       300001    410000
     CR382132_04       400001    510000
     CR382132_05       500001    610000
     CR382132_06       600001    710000
     CR382132_07       700001    810000
     CR382132_08       800001    910000
     CR382132_09       900001   1010000
     CR382132_10      1000001   1110000
     CR382132_11      1100001   1210000
     CR382132_12      1200001   1310000
     CR382132_13      1300001   1410000
     CR382132_14      1400001   1510000
     CR382132_15      1500001   1610000
     CR382132_16      1600001   1710000
     CR382132_17      1700001   1810000
     CR382132_18      1800001   1910000
     CR382132_19      1900001   2010000
     CR382132_20      2000001   2110000
     CR382132_21      2100001   2210000
     CR382132_22      2200001   2310000
     CR382132_23      2300001   2410000
     CR382132_24      2400001   2510000
     CR382132_25      2500001   2610000
     CR382132_26      2600001   2710000
     CR382132_27      2700001   2810000
     CR382132_28      2800001   2910000
     CR382132_29      2900001   3010000
     CR382132_30      3000001   3110000
     CR382132_31      3100001   3210000
     CR382132_32      3200001   3310000
     CR382132_33      3300001   3410000
     CR382132_34      3400001   3510000
     CR382132_35      3500001   3610000
     CR382132_36      3600001   3710000
     CR382132_37      3700001   3810000
     CR382132_38      3800001   3910000
     CR382132_39      3900001   4003362
Continuation (17 of 40) of CR382132 from base 1600001 (CR382132 Yarrowia lipolytica chro

Alignment Scores:
Pred. No.:              1.99          Length:        110000
Score:                  84.00         Matches:       12
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     15            Gaps:          0

US-09-858-332G-15 (1-12) x CR382132_16 (1-110000)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      25949 CACAACCACAACCACAACCACAACCACAACCACAAC 25914

RESULT 45
AE016820_05/c
WPCOMMENT
  Sequence split into 15 fragments  LOCUS AE016820 Accession AE016820
     Fragment Name      Begin     End
     AE016820_00            1    110000
     AE016820_01       100001    210000
     AE016820_02       200001    310000
     AE016820_03       300001    410000
     AE016820_04       400001    510000
     AE016820_05       500001    610000
     AE016820_06       600001    710000
     AE016820_07       700001    810000
     AE016820_08       800001    910000
     AE016820_09       900001   1010000
     AE016820_10      1000001   1110000
     AE016820_11      1100001   1210000
     AE016820_12      1200001   1310000
     AE016820_13      1300001   1410000
     AE016820_14      1400001   1476513
Continuation (6 of 15) of AE016820 from base 500001 (AE016820 Ashbya gossypii (= Eremothe

Alignment Scores:
Pred. No.:              1.99          Length:        110000
Score:                  84.00         Matches:       12
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     15            Gaps:          0

US-09-858-332G-15 (1-12) x AE016820_05 (1-110000)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      91897 CACAACCACAACCACAACCACAACCACAACCACAAC 91862

RESULT 46
AE017344_16/c
WPCOMMENT
  Sequence split into 18 fragments  LOCUS AE017344 Accession AE017344
     Fragment Name      Begin     End
     AE017344_00            1    110000
     AE017344_01       100001    210000
     AE017344_02       200001    310000
     AE017344_03       300001    410000
     AE017344_04       400001    510000
     AE017344_05       500001    610000
     AE017344_06       600001    710000
     AE017344_07       700001    810000
     AE017344_08       800001    910000
     AE017344_09       900001   1010000
     AE017344_10      1000001   1110000
     AE017344_11      1100001   1210000
     AE017344_12      1200001   1310000
     AE017344_13      1300001   1410000
     AE017344_14      1400001   1510000
     AE017344_15      1500001   1610000
     AE017344_16      1600001   1710000
     AE017344_17      1700001   1783081
Continuation (17 of 18) of AE017344 from base 1600001 (AE017344 Cryptococcus neoformans

Alignment Scores:
Pred. No.:              1.99          Length:        110000
Score:                  84.00         Matches:       12
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     15            Gaps:          0

US-09-858-332G-15 (1-12) x AE017344_16 (1-110000)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      75334 CACAACCACAACCACAACCACAACCACAACCACAAC 75299

RESULT 47
```

```
AP007372
LOCUS       AP007372              110560 bp    DNA     linear   HTG 28-DEC-2004
DEFINITION  Lotus corniculatus var. japonicus clone LjT37C05, *** SEQUENCING IN
            PROGRESS ***, 29 unordered pieces.
ACCESSION   AP007372
VERSION     AP007372.1  GI:56805688
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1
  AUTHORS   Kaneko,T., Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Structural Analysis of a Lotus japonicus Genome. XI. Sequence
            Features and Mapping of Nine hundred twenty-one TAC Clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 110560)
  AUTHORS   Sato,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-2004) Shusei Sato, Kazusa DNA Research Institute,
            Department of Plant Gene Research; 2-6-7 Kazusa-kamatari, Kisarazu,
            Chiba, 292-0818, Japan (E-mail:ssato@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/, Tel:81-438-52-3935(ex.2337),
            Fax:81-438-52-3934)
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 29 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1834: contig of 1834 bp in length
            *     1835     1934: gap of unknown length
            *     1935     2876: contig of 942 bp in length
            *     2877     2976: gap of unknown length
            *     2977     4001: contig of 1025 bp in length
            *     4002     4101: gap of unknown length
            *     4102     5069: contig of 968 bp in length
            *     5070     5169: gap of unknown length
            *     5170     6230: contig of 1061 bp in length
            *     6231     6330: gap of unknown length
            *     6331     7480: contig of 1150 bp in length
            *     7481     7580: gap of unknown length
            *     7581     8597: contig of 1017 bp in length
            *     8598     8697: gap of unknown length
            *     8698    11163: contig of 2466 bp in length
            *    11164    11263: gap of unknown length
            *    11264    13548: contig of 2285 bp in length
            *    13549    13648: gap of unknown length
            *    13649    15008: contig of 1360 bp in length
            *    15009    15108: gap of unknown length
            *    15109    17237: contig of 2129 bp in length
            *    17238    17337: gap of unknown length
            *    17338    19607: contig of 2270 bp in length
            *    19608    19707: gap of unknown length
            *    19708    21895: contig of 2188 bp in length
            *    21896    21995: gap of unknown length
            *    21996    24256: contig of 2261 bp in length
            *    24257    24356: gap of unknown length
            *    24357    26977: contig of 2621 bp in length
            *    26978    27077: gap of unknown length
            *    27078    29526: contig of 2449 bp in length
            *    29527    29626: gap of unknown length
            *    29627    33037: contig of 3411 bp in length
            *    33038    33137: gap of unknown length
            *    33138    36578: contig of 3441 bp in length
            *    36579    36678: gap of unknown length
            *    36679    40672: contig of 3994 bp in length
            *    40673    40772: gap of unknown length
            *    40773    44784: contig of 4012 bp in length
            *    44785    44884: gap of unknown length
            *    44885    52276: contig of 7392 bp in length
            *    52277    52376: gap of unknown length
            *    52377    60503: contig of 8127 bp in length
            *    60504    60603: gap of unknown length
            *    60604    66268: contig of 5665 bp in length
            *    66269    66368: gap of unknown length
            *    66369    72326: contig of 5958 bp in length
            *    72327    72426: gap of unknown length
            *    72427    78211: contig of 5785 bp in length
            *    78212    78311: gap of unknown length
            *    78312    87878: contig of 9567 bp in length
            *    87879    87978: gap of unknown length
            *    87979    98711: contig of 10733 bp in length
            *    98712    98811: gap of unknown length
            *    98812   109321: contig of 10510 bp in length
            *   109322   109421: gap of unknown length
            *   109422   110560: contig of 1139 bp in length.
FEATURES             Location/Qualifiers
     source          1. .110560
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="genomic DNA"
                     /variety="japonicus"
                     /db_xref="taxon:34305"
                     /clone="LjT37C05"
                     /clone_lib="LjT library"
                     /note="TAC clone:TM0464, synonym:Lotus japonicus"
     gap             1835. .1934
                     /estimated_length=unknown
     gap             2877. .2976
                     /estimated_length=unknown
     gap             4002. .4101
                     /estimated_length=unknown
     gap             5070. .5169
                     /estimated_length=unknown
     gap             6231. .6330
                     /estimated_length=unknown
     gap             7481. .7580
                     /estimated_length=unknown
     gap             8598. .8697
                     /estimated_length=unknown
     gap             11164. .11263
                     /estimated_length=unknown
     gap             13549. .13648
                     /estimated_length=unknown
     gap             15009. .15108
                     /estimated_length=unknown
     gap             17238. .17337
                     /estimated_length=unknown
     gap             19608. .19707
                     /estimated_length=unknown
     gap             21896. .21995
                     /estimated_length=unknown
     gap             24257. .24356
                     /estimated_length=unknown
     gap             26978. .27077
                     /estimated_length=unknown
     gap             29527. .29626
                     /estimated_length=unknown
     gap             33038. .33137
                     /estimated_length=unknown
     gap             36579. .36678
                     /estimated_length=unknown
     gap             40673. .40772
                     /estimated_length=unknown
     gap             44785. .44884
                     /estimated_length=unknown
     gap             52277. .52376
                     /estimated_length=unknown
     gap             60504. .60603
                     /estimated_length=unknown
     gap             66269. .66368
                     /estimated_length=unknown
```

```
     gap             72327. .72426
                     /estimated_length=unknown
     gap             78212. .78311
                     /estimated_length=unknown
     gap             87879. .87978
                     /estimated_length=unknown
     gap             98712. .98811
                     /estimated_length=unknown
     gap             109322. .109421
                     /estimated_length=unknown
ORIGIN
```

```
Alignment Scores:
Pred. No.:                2              Length:         110560
Score:                    84.00          Matches:        12
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       14             Gaps:           0

US-09-858-332G-15 (1-12) x AP007372 (1-110560)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      43029 CATAACCATAATCATAATCATAACCATAACCATAAC 43064
```

```
RESULT 48
AP004065/c
LOCUS       AP004065              112057 bp    DNA     linear   HTG 21-MAR-2002
DEFINITION  Oryza sativa (japonica cultivar-group) chromosome 8 clone
            OJ1521_B01, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AP004065
VERSION     AP004065.1  GI:15208433
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 8, BAC
            clone:OJ1521_B01
  JOURNAL   Published Only in Database (2001)
REFERENCE   2  (bases 1 to 112057)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-AUG-2001) Takuji Sasaki, National Institute of
            Agrobiological Resources, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     The nucleotide sequence of this BAC clone was generated by
            combining Monsanto and RGP-Japan sequencing data.
            NOTE: It currently consists of 1 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces is believed
            to be correct as given, however the sizes of the gaps between them
            are based on estimates that have provided by the submitter. This
            sequence will be replaced by the finished sequence as soon as it is
            available and the accession number will be preserved.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .112057
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="8"
                     /clone="OJ1521_B01"
ORIGIN
```

```
Alignment Scores:
Pred. No.:                2.03           Length:         112057
Score:                    84.00          Matches:        12
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       14             Gaps:           0

US-09-858-332G-15 (1-12) x AP004065 (1-112057)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     108177 CACAACCACAACCACAACCACAATCACAACCACAAT 108142
```

```
RESULT 49
AC151482/c
LOCUS       AC151482              122984 bp    DNA     linear   HTG 14-JUL-2005
DEFINITION  Medicago truncatula clone mth2-156b13, WORKING DRAFT SEQUENCE, 8
            unordered pieces.
ACCESSION   AC151482
VERSION     AC151482.16  GI:70794848
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 122984)
  AUTHORS   Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
            and Roe,B.A.
  TITLE     Medicago truncatula BAC Clone mth2-156b13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 122984)
  AUTHORS   Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
            and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-SEP-2004) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 122984)
  AUTHORS   Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
            and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-2005) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Jul 14, 2005 this sequence version replaced gi:62734780.
            -------------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ----------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2369: contig of 2369 bp in length
            *     2370     2469: gap of unknown length
            *     2470     7372: contig of 4903 bp in length
            *     7373     7472: gap of unknown length
            *     7473    16926: contig of 9454 bp in length
            *    16927    17026: gap of unknown length
            *    17027    24777: contig of 7751 bp in length
            *    24778    24877: gap of unknown length
            *    24878    44320: contig of 19443 bp in length
            *    44321    44420: gap of unknown length
            *    44421    59635: contig of 15215 bp in length
```

```
            *    59636    59735: gap of unknown length
            *    59736    85221: contig of 25486 bp in length
            *    85222    85321: gap of unknown length
            *    85322   122984: contig of 37663 bp in length.
FEATURES             Location/Qualifiers
     source          1. .122984
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3880"
                     /clone="mth2-156b13"
                     /clone_lib="Medicago truncatula BAC library H2"
     gap             2370. .2469
                     /estimated_length=unknown
     gap             7373. .7472
                     /estimated_length=unknown
     gap             16927. .17026
                     /estimated_length=unknown
     gap             24778. .24877
                     /estimated_length=unknown
     gap             44321. .44420
                     /estimated_length=unknown
     gap             59636. .59735
                     /estimated_length=unknown
     gap             85222. .85321
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:                2.22          Length:          122984
Score:                    84.00         Matches:         12
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       14            Gaps:            0

US-09-858-332G-15 (1-12) x AC151482 (1-122984)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      74474 CATAATCATAATCATAATCATAATCATAATCATAAT 74439

RESULT 50
AC157473/c
LOCUS       AC157473              125310 bp    DNA     linear   PLN 16-FEB-2005
DEFINITION  Medicago truncatula chromosome 7 BAC clone mte1-11e4, complete
            sequence.
ACCESSION   AC157473
VERSION     AC157473.1  GI:59858724
KEYWORDS    HTG.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 125310)
  AUTHORS   Town,C.D., Tallon,L.J., Arbogast,T., Althoff,R., Hine,E.,
            Monaghan,E., Smith,S.A., Utterback,T., Feldblyum,T., Koo,H. and
            Cheung,F.
  TITLE     Medicago truncatula BAC genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 125310)
  AUTHORS   Town,C.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2005) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
COMMENT     Exceptions Reported =   Region (8172-8198) with a mononucleotide
            (T) repeat of 27 to 28 nucleotides.  Region (27247-27300) with
            dinucleotide (AT) repeat of 25 to 27 copies.
FEATURES             Location/Qualifiers
     source          1. .125310
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3880"
                     /chromosome="7"
                     /clone="mte1-11e4"
ORIGIN

Alignment Scores:
Pred. No.:                2.26          Length:          125310
Score:                    84.00         Matches:         12
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       15            Gaps:            0

US-09-858-332G-15 (1-12) x AC157473 (1-125310)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db       4903 CATAATCATAATCATAATCATAATCATAATCATAAT 4868

RESULT 51
AC115599
LOCUS       AC115599              125623 bp    DNA     linear   INV 12-MAR-2003
DEFINITION  Dictyostelium discoideum chromosome 2 map 4229098-4354721 strain
            AX4, complete sequence.
ACCESSION   AC115599
VERSION     AC115599.2  GI:28828893
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 125623)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 125623)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 125623)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 125623)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On Mar 4, 2003 this sequence version replaced gi:19570032.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES             Location/Qualifiers
     source          1. .125623
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /chromosome="2"
                     /map="4229098-4354721"
     CDS             join(<416. .688,837. .1349,1434. .1493,1638. .2931)
                     /note="GeneID exon scores (in order of location ranges):
```

10.12, 21.67, 1.80, 118.27 - GSCJ_ID dd_00183"
/codon_start=2
/product="similar to Dictyostelium discoideum (Slime mold). Gp138"
/protein_id="AAL92328.2"
/db_xref="GI:28828944"
/translation="FKGNQKQQVNALLSQIYQGNVYEGFQCMLNMVMCSGDTVIGINI FTGVPVTLTVDFSVFEDLTYFISSSAISFSPNIFSRLPNFTSPEVYFNIGKMNDQIPE DIVLPENLYTVEFKSISVPLPNAFFSTGAIGTILINNANPGFALPNQFPQNNFLKQLN LNIIGNGFFPSNAGSALSNLTELTLNINNYNTKNITFPTFEPFNQLTKLAINFHSADP LNQLFEFEPPINNINPLEELHIENLGFNISNDVDLTNLSKNLKLTFSNHILPLNKFKY PLNCSLDLSNYPVSIYDVDFRNLTSLSLSNVNYGENLPPASNFDFEKLTYLLLSSNQF NGSLPEEYCSFVKGSLALGSNDLTGSVPDCFKCLGGEEFPGLFPNNFDDFFDDTPADV CTAFSLDVPVNQLLKTNETTVFEFTGSNLGWDSDIESTQANVELAIIEPNKKIKITIP PGAGIQNATLKFGSNFSIEVILNYEFYGPTIDSYYVQGSGIFFRGSYFSYNQDYENIF TINKDYKFTAHIQTIEDQNLNNGVGFENDDTPISVLKNREAFNVSIDVAGKQSETVTF IYFGNITMLSDLSSIQLNTTGGDFSFDGEFGCNTTDFSNPIFKINDIDVQILEITETR INITYPPIENAGTYSLYIEVGDFSFTNEIKFIETPTFPPTPTPTSTPTPTPTSTSTST SDDEGLSTSPTLSISFIYLFSIILLINFVF"
CDS complement(join(3950. .4537,4680. .4808))
/note="GeneID exon scores (in order of location ranges): 38.84, 6.50 - GSCJ_ID dd_00185"
/codon_start=1
/product="similar to Ralstonia solanacearum (Pseudomonas solanacearum). Putative GST-related protein (EC 2.5.1.18)"
/protein_id="AAO51485.1"
/db_xref="GI:28828894"
/translation="MQNNKFDFTIINDHKSPNTYKVELMLIELGVSYESISINDYLKN NSNNRIPALIDNTFKVPNILYESGSVLIYLANKFSKLLPDFRLNATLQNNNSQIISKS VWQLVNLIPSFNNNNNNNNNNNNNNNNNNNNENEVKKVLQLLDDRLSTNQYLGGSQYSI ADISNSVWLLHLLKNSNSNQIISRFSNINKWLELINQRDSIKKLNNLMNNNFNNNSLN NLNNNGIKFSNDLGIAINAL"
CDS complement(5354. .7216)
/note="GeneID exon scores (in order of location ranges): 130.35 - GSCJ_ID dd_00186"
/codon_start=1
/product="similar to Plasmodium falciparum (isolate 3D7). Hypothetical protein"
/protein_id="AAO51486.1"
/db_xref="GI:28828895"
/translation="MGFITEWEYNNRKKDLGVDDDNNNNSINNKTTSNDSNDYYNNNN YSSSSYNNNNNNDYYSNNDYNQIIPVIQNDYQNDSTTNNNNYSNNNNDYYNQVIPVLN NDFQTDTTNNSNTYDYYNNYNYENNNNNNSNTYDYSNNNNLNDNYNNNSFYQQEDYSN NQINNSNNSSNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNLNENIEKDDDIYVIKIKDPI IVKKNSNLIGNKNIDNEKSNKVKEYPLKVKDSYTIHKDIGNNIVLYFKDKACNYQSDL IAIKDKSKTDILKEIYKFVGLDYEKSISTLLIFKMIINDKRTAKPNEFYNYSPPSETI LIPIEFSDPELLKPNHLLIQFGSFPIHRNRWNPTTKFTSTYNFTGTTEYFKGYKGDPV KIDRVKCPYVIDKFKSDFSIPQYGFTKYMGVPSEQITNFYSPSIRPEPCKPIVYHNDN NNSERNATWLKLNFLILFRYLQSRAVNDGRTVTYGEQLPNLVSKDSSIEILEPLKDSI NQSKCLICSRSTLSFGNFLCNTHYSKDLMDAFEKCKPYLFLDANNEPIKQLPKDQLES FKENEKKKEKTVFIKDKFAYIDPSNLQSILDKKVKPYTQENNNYNYNSNYDDYY"
CDS complement(7921. .9687)
/note="GeneID exon scores (in order of location ranges): 222.87 - GSCJ_ID dd_00188"
/codon_start=1
/product="similar to Agrobacterium tumefaciens (strain C58 / ATCC 33970). 3-methylcrotonoyl-CoA carboxylase beta subunit"
/protein_id="AAL92331.2"
/db_xref="GI:28828896"
/translation="MLKSISLLKNNQILLKNIINNGRIINNVGEKLSSKSLLKINYSS STTDRTFNILDGTIDKNSAEYKDNLINMNSTLKQLKENIEKIKLGGGEKLNQKNISRG KLLVRERIEALIDVGSPFLEFSQLAGWGMYGKEEVAAGGIITGIGKIHGVECVIVAND STVKGGTYFPITVKKHLRAQEIAQENNLPCIYLVDSGGANLPRQADVFPDRDHFGRIF FNQANMSAKRIPQIAVVMGSCTAGGAYVPAMADESVIVKGTGTIFLGGPPLVKAATGE IVTSEELGGADLHCRTSGVTDHYARDDAEAIAITRRIVSNLNRKKQPSPVITETEEPL YPTSELAGIVPSDLKKNFDIRKVIARLVDGSRFDEFKELYGTTLICGFARVHGMPVGI IANNGILFSESAVKGAHFIELCNQRGIPLVFLQNITGFMVGKTYESKGIAKDGAKMVM AVATAKVPKITMIIGGSFGAGNYGMCGRSYSPRFLYMWPNAKISVMGGEQAASVLAQI QKDNMAKENKQWSPEEENTFKKPISDKFEEEGSIYYSSARCWDDGVIDPQDSRKVIAL SLSACMNQPINPPSDGFGVFRM"
CDS complement(9983. .11383)
/note="GeneID exon scores (in order of location ranges):

98.71 - GSCJ_ID dd_00191"
/codon_start=1
/product="similar to conserved hypothetical protein [Schizosaccharomyces pombe]"
/protein_id="AAL92333.2"
/db_xref="GI:28828897"
/translation="MRATHSRSVDVKTESTINIKTESTTKVKTESAIDRFKSKSKKTK KHHKVVKSNSDNDNDDHSNINNERKKVKKSITSLKNKKKPSKIKKDLEPVVDQRQRDT AKRAKELKYLEDNRVLEGENFYENVKLLETQTKKNYKLNGEILRKLSVDINKLSERVR CYRDRTIILKRLEEVIKRETSLNKFGEIKVEIFGSSSTQLALKKSDVDIVMSFETELT KRNDITKWCYQFSSILRANGFYNIKPIIHAKVPIVKFFDPKTEFHCDITLTKDSGNTG VVKEFCQLLPILPVLIIFCKNWASVLNINDASQGTLSSYSITNMVIFVLQKKGLLPSY KDIQSFKNSSKEVQYVVDNRDIADLLVYFFKYFGYIFDYRNSMVNILDYNNDKKEPMD AEEELKFQNNREYDNKNHVLPLVDMKLGDTKDLFFIKDCIAGHNITRCILRSQLFTLI KYFQKAYLNLISGKLP"
CDS 13531. .13833
/note="GeneID exon scores (in order of location ranges): 16.95 - GSCJ_ID dd_00194"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAL92334.2"
/db_xref="GI:28828898"
/translation="MSIIYSPEVPFQLFEVELRDLDVNSEEFKFVAQIFNVGVPSTIK KIQVVYNKSVWKNYNKNLGNSLKEIFVFHGSRLNEPKLIYTNGLLVEKTSAGLYFA"
CDS 14573. .14956
/note="GeneID exon scores (in order of location ranges): 6.39 - GSCJ_ID dd_00176"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAL92335.2"
/db_xref="GI:28828899"
/translation="MNSFFVKKIQVVYNKKVWDNYYQHWLKSNHSLEELFVFHGTSLN DPSLIYTNGLRAEKTQSGLYFAIKSESNGFTYKNTDCSQIFICRILIPRQNVSPRFHV IKNNDHHYPQYLISYNSKYRNSIML"
CDS join(15673. .16125,16226. .17482)
/note="GeneID exon scores (in order of location ranges): 49.65, 142.60 - GSCJ_ID dd_00179"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAL92336.2"
/db_xref="GI:28828900"
/translation="MNKILSILLLSILILFILRVESYPTYGGSNSANIRREDRHSFRA PFLQSGTGILYWDIEGSTIVNDDFIRLTSDQKSLHGAIWNTEPMEQPWWEVVFEFRVH GAGRIGADGIALWLVDRKEGNSQDFSIYGSKNLWKGLAIIFDTFDNDQNGDHPLISVF YNDGTKFYETAKDGSNMKLGSCSSRYRNDKHNAKSRIRYYHGLLSVEIDPNGSGIFEK CVQDVRLDIPTRYTFGVSAATGGLTDNHDVYSFDTFSLDASVSNQNTNNYNNVQFERE RLEKQRLEYELQQTQQQQTQQQQQQQQQQQTQQQTQQNNQDKEYQEFLQYKAQQQQQ NNYQQTQNDLIEKLNEIQQEQQQTQPPQQQQQQTQPPQQKLQDEIKNLSEKEQIGFLV KKVDSLVEAIHSLHVILDDPDYKKSFDGLRDQVNVISQALVSLSQGSLKVKDFESFRS TYDQKQSQFIKDFYDFKSTMQDLKATINNRFNNQDNDVTKSIESVNMNLNSLRKLVEA STRDTQKISSILNSNADELANTIQKHSSFGFWTYFLFVQAIFIFGFVFWRKYKDDNNK KLI"
CDS 18511. .20217
/note="GeneID exon scores (in order of location ranges): 169.96 - GSCJ_ID dd_00184"
/codon_start=1
/product="similar to Erwinia amylovora. Amylovoran biosynthesis protein amsF precursor"
/protein_id="AAO51487.1"
/db_xref="GI:28828901"
/translation="MSVIINSIADLKTYEPTSTVTTVYVKAFTVNTNIGGGDFYYDSA DTTTVTDNATVIVTNGGKRWKKVISDFNDLNVTHFGALKDGVTDCSAACITMHNWSVK YAPKLGIRFPAGTFKISGIDLSNTSIAHFRMTGSTVAYGYFPATTLISDQGTGMMVKV KARWTEISNLIIYGENDKVVNTKGFYQNTEIEGEFIRITNINFRNLGGVCVSMVDTLD TKIDQFYVTRCTGGVIYGTWSGSATGSWDHLTAVELTNFNIQNCTGANIFDLQRCTQS IIRNGWIEHSDSPGDLSNGQWVIEALSMEDCVNPLNLTFCRYIITQKNLQGTSAILTN DSTKSTWLSIWENGGTEIENYGIRTTGSMNYGQLTSEFRFSNSSASSQWLRLGEVFIP SDGDTLNIRIVGSLGFSSLTNSHDVASGRHGNGEALLRVQKKSPGVQMSWEGRGSCPV TDVMYLPATNKTNCSIYIKLGPYVMNTVAILETSARDRYYAGVCFRWMYDGTKLDETA VTTIEGIVQALPQASWNAGKNGIAIGSDGYFGITTVPIGTDNQLPIYINGSLYKIAVV KQ"
CDS 21628. .21846
/note="GeneID exon scores (in order of location ranges):

```
                    6.32 - GSCJ_ID dd_00190"
                    /codon_start=1

Alignment Scores:
Pred. No.:              2.27          Length:         125623
Score:                  84.00         Matches:        12
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     2             Gaps:           0

US-09-858-332G-15 (1-12) x AC115599 (1-125623)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      47074 CACAACCACAACCACAACCACAACCACAACCACAAC 47109

RESULT 52
AC115592/c
LOCUS       AC115592               125958 bp    DNA     linear   INV 11-MAR-2003
DEFINITION  Dictyostelium discoideum chromosome 2 map 1-125959 strain AX4,
            complete sequence.
ACCESSION   AC115592 AC115583 AC116029 AC117077
VERSION     AC115592.2  GI:28829078
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 125958)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 125958)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 125958)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 125958)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On or before Mar 4, 2003 this sequence version replaced
            gi:20066250, gi:19747091, gi:19569925, gi:19569975.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES              Location/Qualifiers
     source           1. .125958
                      /organism="Dictyostelium discoideum"
                      /mol_type="genomic DNA"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /chromosome="2"
                      /map="1-125959"
     CDS              join(5527. .5782,5865. .6021,6132. .8390,8502. .8802)
                      /note="GeneID exon scores (in order of location ranges):
                      28.78, 10.55, 225.85, 35.01 - GSCJ_ID dd_01188"
                      /codon_start=1
                      /product="similar to Dictyostelium discoideum (Slime
                      mold). Kinesin 1 (Fragment)"
                      /protein_id="AAM09366.1"
                      /db_xref="GI:20066259"
                      /translation="MATSCNIRVMCRFRPLNEREKALKENQTCVTFPDETQVIVSGQP
                      FTFDRVFTPESTQKEVFESVKDTIHDVLLGYNGTLLAYGQTGSGKTFTMGSAAAESDF
                      ENVEQLGIIPRGNHLIFNTIAEESDGNAEFTIKCSYLEIYMENIQDLLNPKNNKQLKI
                      RESKSQGIYVEGLTEEYVASEEDIMELIQVGESSRSVAKTNMNQRSSRSHSILIIAIE
                      QKSSDGSKKRGKLNLVDLAGSEKVSKTGAEGIVLEQAKKINQSLSLLGNCIHALTDSK
                      REHIPFRDSKLTRLLQDSLGGNTKTTLLVTASPHFNNVDETISTLKFGARAKSIKNNV
                      KVNQEKSAAELQIIVNALTKELSILKVYSISLENLVNYFKSSSYQPGNPIPKELEPNK
                      QNLLLLQQQSNSSSGGGGSGSSGGSSNGSLMMKPRSTTPTPPSINRPHSSASTHRHSI
                      AITGTHSKEGSGGGLTSSISSSSISSMSSLSSSIDNDYNGSSLDDSNGSNGLFNPLAI
                      VEMSIEMEKMKEDTQLLIEKFKDEISEITIQYQSTQEELNQCRQQLDQIKEQLEQQRS
                      QFIKEQSLLKESERNATLDSTSKDLKIQSLISKIEDLRLLASQVIQYLERKRLSDDFD
                      IGIFMGSQDGANGDSANMFSLISNSIDQGTYEDDVNIEDIIRYLSEEEVLTMQVKLQL
                      QNKVHQLEQKIQQLVSDLNTTEINYNQSILQCQKFESENSLIKRKFKSMFSSSNNNNN
                      NNNSPPSSPSSKLLIQSSNNNNNFDSNLNSSSLSSSLPSSNGIEQDQQVDHQVDHQVE
                      NDHLVNDEDNKLLDENSKLKENELKLLLEIKNLKLQSEKSNDETLKWKDELSIKSALY
                      QNQILNLQNENQSLSNKLNVEKQQKQSSQSQQIEFATKLNDLIKSSEDDKELYRNEKN
                      QMELEIASLKASLDEMDLKNKELQDQLISTQRLIGARRVVKIVRGGADSMKTALATKE
                      VFGQFTLRKTENSKTLFQ"
CDS                   9421. .13035
                      /note="GeneID exon scores (in order of location ranges):
                      338.45 - GSCJ_ID dd_01189"
                      /codon_start=1
                      /product="similar to Dictyostelium discoideum (Slime
                      mold). TRFA"
                      /protein_id="AAM09367.1"
                      /db_xref="GI:20066260"
                      /translation="MNRKSTLNFTNSSPVGTRRNTPSNNNSNNFNYNNNNSNNNNNNN
                      NNFNSNNNNYNNNNFNNNNFNNNNNNNNYNHNNNNNQSQVSLNNNHQRNTGNQNLQN
                      NQQPQQQPQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQIQ
                      NYNYNRIDWRYLLSVDIDNIVKSNDVDTLRSVIQILLGISLIHIENDQKFKNLDEGIL
                      KVVRLSQLAIQYLQDMNDSYEDDLKVLNESNSFYTHQIEHLKAKLLLQSDELSFINRE
                      RIFELSIKAEKLQYQLQQQQLNDQNLNIHNNNNNNDNSSLFLLNNNPNSSIYQNNYNK
                      FNANANNNNNNTNNNKQIENGNNTYQQQQQQQQQQQQQPNNNNNNTNNNNSNSQNVIA
                      NRSNNNDVYNNSGGGNLQQNGSIRNHLEEIAMLNAKNNKLESLLVLVQNDFINYKEEL
                      SVRDESKSKQSERERKKITQLTKKLETQAEKIEKLEEFINSNFQNQQQYDYNDQQDYI
                      DQQQQQQFLLDQQYMYEQQQIEQQQQIEQQQFEQQIEQQQIEQQQQIEQQQQQFEQQQ
                      FEQQQQFEQQQQLQKQQQQHQIQFQGRQFEQQQQQQQQQQQQQQQQQQQQQQQQQQQQ
                      QLQQQFQPKPKTPQKFRLNSPQLNQISPQKQRQQQIEQLEFQLQQRQDRENEEYDQELI
                      KEQSLQQQQQQQRNNDSEDDEYDDDYQTQQQKQQQFNNHQNFQKPLKLNVNNNNNNNN
                      NNNNNNNNNNNNNNNNNNNNNNNNNSSSSNSSNNNFSGNDSSSGSSSSSGSSGSSNNN
                      KNNKNNNSNNSNNNSNNNSSEYEDQDQFQDQYHGQDQGQLQDQDQEQGQEQEQEQIEN
                      LPERQPQPISLLNRPYKLQSFTTKHSDESIELERENIINIMNQRLTEKGISIEKGISK
                      KNLSQFLNQLEIERQYDSNYLDQKEMRDYLESYVDDLLASQLFSIVENFNNQQNNNNN
                      NNNNSNNNNNNNNNSNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
                      NNNNNNNNKNNRTNNKNQVKIPKEESPLNLSSGNTSPIDTQFNQQQQQQQQQKLKNR
                      NKLIAHKDKTASFDVSSDENIDDQDILNNSFNFNNFNNNDNNNINNNNNNNNINNNIN
                      NNNNNNNNNNNNNNNNNNNNNNNNNNNSTKNNNFNVQSQNSSPNSPTTPRFILNYQNLI
                      "
CDS                   complement(13390. .15816)
                      /note="GeneID exon scores (in order of location ranges):
                      186.41 - GSCJ_ID dd_02368"
                      /codon_start=1
                      /product="similar to Dictyostelium discoideum (Slime
                      mold). Tortoise"
                      /protein_id="AAM09368.1"
                      /db_xref="GI:20066261"
                      /translation="MNKTIIINNVSKYIKFKNYYSSSSTILKKKTSEIYKIKGLDIKD
                      RKELYSLNNDSIKKKLNQLKDETNQLLKERGEELMKDLSKTLNLTSDNLKYNITKSPF
                      TLRAYLLAANQGISFYHHELRKQINPTDDYERQKRKDLVYKLQSQEIDSLTSGGANKK
                      KSPFLEDNNNKKSMSIEREMEIIREQQQEQQHFDQQMSNSTKWEMSQLEIQPRIDTFR
                      QDELFVSFNPYESMSTNSQSYQLVSRLSKFVWNKELTNWEMFSSSKLNQLIPQLQLHI
                      LDKINLNNDNDNDNSIIINQFLESIKFISNEINSTDLNIFKTLNSSNNNNNSSENIIE
                      VENGGDFKKIINNSIVDSLIYMKGHHFRITSDPFILFFQQLKENQDYFTSNEDYKQFL
                      EKLLDTLLFETIEFDPTSTTTINGNRQFEYKLKLDLLQREAFRQFKVYQVFNTSPDAK
                      RHEESDLRKLAHLCKTLQKRSTSDFIQESKKYLSPQILSNGLLNSNKLLVATSSSNEL
                      NQSKTSISISQLTDGLHSCIPATMSLLEFSIKNSSKDNILFDTFDGKLKNINSLDDLI
                      TKFANSKYIYLRESLSQRFSMFIQSSNNTTDNNTTTDTNDNDIIIGLNNPIYKLLIDL
                      EDKVLELSQNNFEENPIENWIGCLSIELGDEDDNIEYKDTDQFNVIASQNFNLVRINN
```

QMLAQLHKTFSEDLDFQIEELELMIKNKKILEREIKAHNEKISKIIKDSRDSTTPTFY
FIGIQDDQVYTLEIPNEWEPVWNLLKSYPNGLLLKDFIDKQTFVTKEAFLSTLTDLNK
YGVVGLFPKN"

CDS join(16509. .16729,16888. .17250,17350. .18556)
/note="GeneID exon scores (in order of location ranges): 15.53, 15.13, 100.07 - GSCJ_ID dd_00465"
/codon_start=1
/product="similar to Sulfolobus tokodaii. Putative acetyl-CoA synthetase"
/protein_id="AAM09358.1"
/db_xref="GI:20066251"
/translation="MENNKPYKFIDRNLNIIKKSGSYENIYKSFKYEIPEYFNIGLEV
CDYYANDPNKMNNIAIKHIYDTDDKQILKEYTFKEIKIKSNQLANKFKEIGLNKKGEV
VGVFLTQGIECALSHITILRSGFIGLLLSVLFGPEALEYRISSSSASCVITDLDNIHK
LLKVVPNLPSLKKIIVFGDNLNSINYSKHKNIVELWNENICNNYSKEFEPIITKSNEP
AYLIYTSGTTGLPKGCLHAHRVLIGNNVGLQFTLNLFPQNTVSVTENCYYSPADWAWI
GGLLVVFLPSLYFGVPLLAYKSKGPFDAKKLLTLVKTHNVTMSFLPPSALKMIKFQKD
QLKDHKINNMLAINSGGESLGENLLVWSKDQFGVEVAEFIGQSEGNFITGNCLDLFPP
KNGSMGKPIPGSNVRILNDNGERVPHDEVGYLCIEFTGDNANQFLGYWNDPSSTEKRM
IFSKESGTKWIKTGDLAKEDSDGYFWYIGRDDDVINSSGYRIGPNEIEGCLLKHPLIL
NVAVIGIPDEIRGEVVKAYIVLNQSVTPSQEIKKEIQNYVKTQLSAHQYPREIEFINP
NEMPITTTGKIMRNSLRQLHKMKSNNKSKI"

CDS complement(join(18805. .18809,18907. .19296,19808. .19919))
/note="GeneID exon scores (in order of location ranges): -6.01, 14.11, 6.40 - GSCJ_ID dd_00466"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAM09369.1"
/db_xref="GI:20066262"
/translation="MKPKRSLTHLIEHTERIYNALEYGVRIGNISRAAREYGFLLELR
RSKAVGPIGEPVYVTNSTSISETYNVFFTISLDDCLYKIRDKTNNTEDFFQYVRELLE
SGFCKPGSLLVMDNAPIHGGIEALNALFELFNQYSVQLIFLPSYSPELNPIERFFGYI
KNHFYREP"

CDS complement(20933. .22594)
/note="GeneID exon scores (in order of location ranges): 218.64 - GSCJ_ID dd_00467"
/codon_start=1
/product="similar to propionyl-CoA carboxylase beta [Caenorhabditis elegans]"
/protein_id="AAM09359.1"
/db_xref="GI:20066252"
/translation="MLSRRLLNKNTIQFLNKNCKNVSSQGIKSFATGKPTPAKEPFKA
VKYESTIERINGRIQEETERAQLGGGVKKIDVQHGKGKLLARERLDLLLDDGSFREYD
MLVTHRCTDFNVEKTPGDGVVTGHGTINGKLVFVFSQDSTVNGGSLSEAHAEKICKIM
DKAMLVGAPVIGLNDSGGARIQEGIASLAGYAEVFQRNVMASGVIPQISVIMGPCAGG
AVYSPAITDYVFMVKDTSYLFVTGPDVVKSVTNQEISQEGLGGAKIHTTKSGVAHLAF
DNDIDALKRVREFVTLLPSNPNEPAMREAFDSPSREDDVLDQIIPDDPNKPYDMKEVI
ERIVDEGEFFEIQPEFAKNIIVGYARMEGRTVGFVANQPKELAGCLDIDASVKAARFV
RFCDCFNIPIVTLIDVPGFLPGTNQEHNGIIRHGAKLLYAYAEATVPKISIITRKAYG
GAYDVMSSKHLRGDTNYSWPTGQIAVMGSKGAVEIIFRGKGDIKEQEQIYNDKFANPL
PAARRGFIDDIIPPRMSRQYICEDLEILKNKKLSNPTKKHGNIVL"

CDS join(23169. .23191,23654. .24101)
/note="GeneID exon scores (in order of location ranges): -2.82, 13.69 - GSCJ_ID dd_00468"
/pseudo
/codon_start=1

CDS join(24420. .24972,25168. .25328)
/note="GeneID exon scores (in order of location ranges): 33.75, 3.72 - GSCJ_ID dd_00469"
/pseudo
/codon_start=1

CDS join(27215. .27684,27767. .28376)
/note="GeneID exon scores (in order of location ranges): 24.08, 30.66 - GSCJ_ID dd_00470"
/codon_start=1
/product="similar to Dictyostelium discoideum (Slime mold). Guanine nucleotide-binding protein alpha-5 subunit (G alpha 5)"
/protein_id="AAM09370.1"
/db_xref="GI:20066263"
/translation="MGSQFSVLNRKWLIERSIMIEKRKRRSNKLIKILMMGNENSAKS
TFAKKVKSIYQSDKIENSDANSILPYIKLYLSNSFKDVIKFIQQHPQSPDTKPMVSTQ
LGIIAWDKLSSFNYIPFNFKPTKELAKYIYDICHDPLFKSYIYPIISSKREDAFYLIE

Alignment Scores:

Pred. No.:	2.27	Length:	125958
Score:	84.00	Matches:	12
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-15 (1-12) x AC115592 (1-125958)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      68396 CACAATCACAACCACAACCACAACCACAACCACAAC 68361
```

RESULT 53
AP005406/c
LOCUS AP005406 127902 bp DNA linear PLN 03-DEC-2003
DEFINITION Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 8, BAC clone:B1147B12.
ACCESSION AP005406
VERSION AP005406.3 GI:38637280
KEYWORDS .
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1
AUTHORS Sasaki,T., Matsumoto,T. and Katayose,Y.
TITLE Oryza sativa nipponbare(GA3) genomic DNA, chromosome 8, BAC clone:B1147B12
JOURNAL Published Only in Database (2002)
REFERENCE 2 (bases 1 to 127902)
AUTHORS Sasaki,T., Matsumoto,T. and Katayose,Y.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-2002) Takuji Sasaki, National Institute of Agrobiological Sciences, Rice Genome Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/, Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT On Dec 2, 2003 this sequence version replaced gi:33468826.
Genes were predicted from the integrated results of the following: GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH (http://www.softberry.com/), GeneMark.hmm (http://opal.biology.gatech.edu/GeneMark/), GlimmerM (http://www.tigr.org/tdb/glimmerm/glmr_form.html), RiceHMM (http://rgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor (http://bioinformatics.iastate.edu/cgi-bin/sp.cgi), sim4 (http://globin.cse.psu.edu/html/docs/sim4.html), gap2 (http://www.tigr.org/software/glimmerm/), BLASTN and BLASTX. The genomic sequence was searched against NCBI NonRedundant Protein database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at RGP or DDBJ. Protein homologies of the coding regions were searched against NCBI NonRedundant Protein database with BLASTP. ESTs represent the identified cDNA sequences using BLASTN with the corresponding DDBJ accession no. and RGP clone ID. Full-length cDNAs represent the identified cDNA sequences using BLASTN with the corresponding DDBJ accession no.
A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with full-length cDNA or EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted by two or more gene prediction programs is classified as a 'hypothetical' protein according to IRGSP standard. A gene predicted by a single gene prediction program is also classified as a probable 'hypothetical' protein and is included as a miscellaneous feature of the sequence.
The orientation of the sequence is from -21M13 to M13rev of the BAC clone. This sequence of B1147B12 clone has an overlap with P0015C07 (DDBJ: AP004654) clone at 3' end. Detailed information on overlap

```
          and assembly quality together with annotation of this entry is
          available at http://rgp.dna.affrc.go.jp/GenomeSeq.html.
FEATURES             Location/Qualifiers
     source          1. .127902
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="8"
                     /clone="B1147B12"
     gene            2332. .2511
                     /gene="B1147B12.1"
     misc_feature    2332. .2511
                     /gene="B1147B12.1"
                     /note="hypothetical ORF
                     predicted by GlimmerM
                     this category is not included in IRGSP standard"
     gene            3839. .7735
                     /gene="B1147B12.2"
     mRNA            join(<3839. .3868,6548. .6742,7643. .>7735)
                     /gene="B1147B12.2"
                     /note="start and end point are not identified"
     CDS             join(3839. .3868,6548. .6742,7643. .7735)
                     /gene="B1147B12.2"
                     /note="predicted by FGENESH etc."
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="BAD03544.1"
                     /db_xref="GI:38637281"
                     /translation="METDAAGGMELVPNLISGPMIANSIKLFLAQCVDQDIHVLGIEL
                     EIAMVNIHQVIVGDSKAYINDAPNVEARSLKSDKRAQVCLYESVDEKSVLSRWCQICV
                     GVN"
     gene            complement(join(8627. .8640,9382. .9529))
                     /gene="B1147B12.3"
     misc_feature    complement(join(8627. .8640,9382. .9529))
                     /gene="B1147B12.3"
                     /note="hypothetical ORF
                     predicted by GENSCAN
                     this category is not included in IRGSP standard"
     gene            complement(14133. .17716)
                     /gene="B1147B12.4"
     mRNA            complement(join(<14133. .14199,14473. .14590,14671. .14727,
                     14943. .15010,15104. .15144,15405. .15461,15980. .16041,
                     16833. .16929,17016. .17102,17171. .17263,17394. .17452,
                     17653. .>17716))
                     /gene="B1147B12.4"
                     /note="start and end point are not identified"
     CDS             complement(join(14133. .14199,14473. .14590,14671. .14727,
                     14943. .15010,15104. .15144,15405. .15461,15980. .16041,
                     16833. .16929,17016. .17102,17171. .17263,17394. .17452,
                     17653. .17716))
                     /gene="B1147B12.4"
                     /codon_start=1
                     /product="putative Zinc transporter zupT"
                     /protein_id="BAD03545.1"
                     /db_xref="GI:38637282"
                     /translation="MDPKVVVALALSLIGGLSTSLGALLAILNCAPNNKTLGMLQGFA
                     TGLMLSMSFFDLAYDAVNAIGFLKGNLWFFAGALLFSTIAEVFPEPDCNLADENDKHL
                     ENKTGNNNIAGKELMMRHRRRVIFSVVVTAIVAGVSLQNFPVGTAAFLGTTKGFRVGL
                     NLSLAIALHYIPEGISVALPAYFATCSKWQAFKLATLSGFAEPLGVIIVAYLFPSNLN
                     PEILEGLLGLVGGVMAFLTLYEMLPIAFEYAGRKDAVKAVFVGMAFMSMRHERMRMDT
                     YGLVPHHLAALQQ"
     gene            18005. .19493
                     /gene="B1147B12.5"
     mRNA            join(<18005. .18732,18835. .19026,19132. .>19493)
                     /gene="B1147B12.5"
                     /note="supported by full-length cDNA(s): AK108701"
     misc_feature    18005. .19493
                     /gene="B1147B12.5"
                     /note="contains full-length cDNA(s): AK108701
                     non-coding transcript
                     probably inactive due to long 5'UTR in CDS"
     gene            complement(19926. .27167)
                     /gene="B1147B12.6"
     misc_feature    complement(19926. .27167)
                     /gene="B1147B12.6"
                     /note="contains full-length cDNA(s): AK121153,AK102424
                     non-coding transcript
                     probably inactive due to long 3'5'UTR in CDS"
     gene            join(29795. .30195,30289. .30365,31668. .33560,33643. .34147,
                     34250. .34358,34436. .34548,34619. .34806,35481. .36129,
                     36214. .36691,36760. .36855,37042. .37141,37301. .37360,
                     38970. .39282)
                     /gene="B1147B12.7-1"
     mRNA            join(29795. .30195,30289. .30365,31668. .33560,33643. .34147,
                     34250. .34358,34436. .34548,34619. .34806,35481. .36129,
                     36214. .36691,36760. .36855,37042. .37141,37301. .37360,
                     38970. .39282)
                     /gene="B1147B12.7-1"
                     /note="supported by full-length cDNA(s): AK120880"
     CDS             join(30085. .30195,30289. .30365,31668. .33560,33643. .34147,
                     34250. .34358,34436. .34548,34619. .34806,35481. .36129,
                     36214. .36691,36760. .36855,37042. .37109)
                     /gene="B1147B12.7-1"
                     /note="contains full-length cDNA(s): AK120880"
                     /codon_start=1
                     /product="putative PWWP domain protein"
                     /protein_id="BAD03546.1"
                     /db_xref="GI:38637283"
                     /translation="MAPAGRRGAKAHKWTTQPQLGDLVLAKVKGYPPWPAKVSKPEDW
                     DQMPVPRKVFVVFFGTREIAHIALPDLQPFTENAKSEVMDRSRNKQCPKKYIDSFAEA
                     VVEICKAYDDLPKSSETTTCMLPDDQSEQPTEHLVKSPNNDEAPRSGQMEGDSPSDNL
                     NTSGLGSGTEVDIKDGSCDIRDSSLAAVKRKKPKDLDQPKKKKPVTSKSAINMHLEQD
                     CSATTVHAERELEEPKAEKEINPSEFLTLDPTVQIVCALEVPKKSKAAKQSKNAERKD
                     NKRVNVADISGRTTPGAVLDTELKRSAEKDSKGFKKSKLMMKQSVSNESEKIDHKRIM
                     VDKSDKLLARKSSSVFSSNKKPLPGSEQRKLDNSTDMRPAKRPKLMDRANETVKTEAK
                     SETSLHADNEKDSALKIEKSIPAEAVSNSVPKIGVGDDWTRRSGSLLSPLARLHSQGS
                     EPASGSTQLSAVDTAKKVSSLKENFSRVGKPLAKPRRRACRFGDDDEEEQRTPPHKSS
                     VKSIPMRGVPTEKFQSQTGVRGISSSQIGNASAMKFGVAREEKPKSIGRSPVEHEPDY
                     TSPNQDKMHGRLQIMGKRSTTSSVDTSASLGNKTNLADRRSSGQLRMAASSEVKKTQG
                     SSSKLLHQTPGNLHSQNPDDSEKNALLSKSENSKAKTKSGTQITATVEGRISTAMQAE
                     RIGKMDHSKEQRSDFVDKAAFAEPNSDPVKSMKHLIAAAQARRNLIAAAQGKSDGLSA
                     DNTVLSSTPYGLPGLSPGPVFHIPSASRIIPESDGMQFPDSFCAITEPGQQVAMKNLL
                     EIEHEHGKSPKTRQSSDSLSGGTDAAIARDALEGMIETLSRTKDSIGRATRHAIECSK
                     YGIAAEIVELLVLKLENEPNLHRRVDLLFLIDSITQCSHSQKGVAGVSYVPTVQAALP
                     RLLGAAAPPGPGARENRRQCLKVLRLWLERKIMPEGILRRYMDDIEVPNDDANTGFLL
                     RRPSRAERSVDDPIREMEGMLVDEYGSNANFELSGILSSNVFENDDDFPGLSPAISLP
                     VQSGRMQENEQAIAPNFVEENIRLPKDVTSDVPMEDASLLPRDKQQTDGAIPVVHDLQ
                     HEIDREQALADQNELPPLPDGPPPLPLDAPPPPPLPEGPPPLPSDSPPCQPPLPPSPP
                     PATPPPPPPLSPSLPPPPPPPPLPSGPPPQPAPPPLPIQPPPIPPPPVPSSPSSLGYQ
                     PPAPEYFRASNSNQLNQMAGNASIQGIGNMTNFIPGGSVNTQAAVNFTPSVPPDYGNN
                     NLYIKPQGSNGNFQFRPTGVPFQQGTFSAFPSAQTPPVRPHTHLTQMNPVGQQAVPPC
                     NPYAVQSFPNNQSQYTSDEQWRMTSGNFSPDDQRNTWLPGARALSCAEGSFMQDGYPR
                     SNIDRSSMNPMTHQRPVLNRMPSGAPVPGHVPQMLPARPDIHTFNCWRPS"
     gene            join(35492. .36129,36214. .36691,36770. .36855,37042. .37141,
                     37301. .37618)
                     /gene="B1147B12.7-2"
     mRNA            join(35492. .36129,36214. .36691,36770. .36855,37042. .37141,
                     37301. .37618)
                     /gene="B1147B12.7-2"
                     /note="supported by full-length cDNA(s): AK109496"
     CDS             join(35590. .36129,36214. .36691,36770. .36795)
                     /gene="B1147B12.7-2"
                     /note="contains full-length cDNA(s): AK109496"
                     /codon_start=1
                     /product="PWWP domain protein-like"
                     /protein_id="BAD03547.1"
                     /db_xref="GI:38637284"
                     /translation="MQENEQAIAPNFVEENIRLPKDVTSDVPMEDASLLPRDKQQTDG
                     AIPVVHDLQHEIDREQALADQNELPPLPDGPPPLPLDAPPPPPLPEGPPPLPSDSPPC
                     QPPLPPSPPPATPPPPPPLSPSLPPPPPPPPLPSGPPPQPAPPPLPIQPPPIPPPPVP
                     SSPSSLGYQPPAPEYFRASNSNQLNQMAGNASIQGIGNMTNFIPGGSVNTQAAVNFTP
                     SVPPDYGNNNLYIKPQGSNGNFQFRPTGVPFQQGTFSAFPSAQTPPVRPHTHLTQMNP
                     VGQQAVPPCNPYAVQSFPNNQSQYTSDEQWRMTSGNFSPDDQRNTWLPGARALSCAEG

Alignment Scores:
```

```
Pred. No.:              2.31          Length:        127902
Score:                  84.00         Matches:       12
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     15            Gaps:          0

US-09-858-332G-15 (1-12) x AP005406 (1-127902)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      60490 CACAACCACAACCACAACCACAATCACAACCACAAT 60455

RESULT 54
BX465845/c
LOCUS       BX465845              144092 bp    DNA     linear   VRT 29-JAN-2004
DEFINITION  Zebrafish DNA sequence from clone DKEY-261A10 in linkage group 18,
            complete sequence.
ACCESSION   BX465845
VERSION     BX465845.7  GI:41392484
KEYWORDS    HTG.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 144092)
  AUTHORS   Dunn,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JAN-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jan 29, 2004 this sequence version replaced gi:35761513.
            ------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived
            Zebrafish pUC subclones occasionally display inconsistency over the
            length of mononucleotide A/T runs and conserved TA repeats. Where
            this is found the longest good quality representation will be
            submitted.
            Repeat names beginning 'Dr' were identified by the Recon repeat
            discovery system (Zhirong Bao and Sean Eddy, submitted), and those
            beginning 'drr' were identified by Rick Waterman (Stephen Johnson
            lab, WashU).  For further information see
            http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml DKEY-261A10
            is from a Zebrafish BAC library
            VECTOR: pIndigoBAC-5.
FEATURES             Location/Qualifiers
     source          1. .144092
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-261A10"
                     /clone_lib="DanioKey"
ORIGIN

Alignment Scores:
Pred. No.:              2.59          Length:        144092
Score:                  84.00         Matches:       12
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     5             Gaps:          0

US-09-858-332G-15 (1-12) x BX465845 (1-144092)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      83386 CATAATCATAATCATAATCATAATCATAATCATAAT 83351

RESULT 55
AC116551/c
LOCUS       AC116551              153751 bp    DNA     linear   INV 21-FEB-2004
DEFINITION  Dictyostelium discoideum chromosome 2 map complement(581327-427576)
            strain AX4, complete sequence.
ACCESSION   AC116551 AC116101 AC116549 AC116965
VERSION     AC116551.2  GI:28829639
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 153751)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 153751)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 153751)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 153751)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   5  (bases 1 to 153751)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2004) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On or before Feb 21, 2004 this sequence version replaced
            gi:28850309, gi:19807752, gi:19920064, gi:19807784.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES             Location/Qualifiers
     source          1. .153751
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
```

/strain="AX4"
/db_xref="taxon:44689"
/chromosome="2"
/map="complement(581327-427576)"
CDS complement(join(<351. .2554,2687. .2807))
/note="GeneID exon scores (in order of location ranges): 70.76, 0.70 - GSCJ_ID dd_02254"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAO52218.1"
/db_xref="GI:28829702"
/translation="MNESISDHNEKLFWKVFRNKYLNIEIFKQLESTEIEYDSFTIAS
LIKLSVNNSNNNNDNNNKLIEIFKLLMEKEKYLMEGLDLVTISISSRNFEILKILLNK
PYSFLVKETDFNDALQKYGDSEIEKSILQLLGNQININTFLIENNFQIKNSIPFFKNT
IQFILNNQNNDDNNDDNKNNNNKLIKLLELKIYSIKKDEILNLELIEMQNKLLDLNYF
SKENFQNSNIYWFKELYEFKGVKTIFIHLILLVFKIYYKRKSLNIPNEIEIEINEIMN
NSLLIKNEEKEKKLLKLLINETQSNLLFKLYYYRYYEIIEGCKLSKIVYRNLKTFNIF
KLIESNNNSNNNKEEIIRIIKNFINVYEFKNNENDKQLINKFRNVILNNLNSKEYYHK
LIETIGNSQIKTFGLPIEWNINLLDFLLDDEINNKIIIIKDKQVIDWLSNMPRLNNRI
FTFDIKTIYYYKRKKNLIQFQSLELIEYAIKKLETNNNNNSNNDNLPPIKLETINWSM
VYNNPLKSIIDEFKFQNESIINEDELNDKVTFNLFKFLKFEKNEFTILVQLIKYFIEK
VLNKYGQDKSYEILKNFVKIDNLCKFQKETLELLVLSTINLDLIMVMIILNCGIDFTN
KQDEILINFWNVHILTNTSIFYIPTINGNINFFKIITHIKPQYIINNIKLWCEQTSLL
ECKHLAINHLFKKLIQTKGTTLNQILYIKDFINTINNNNNNNNNNNNNNNNIIKLDHS
KFHSIHYLNLKSPKLILYLIESNQIRDWNYLLIPQ"
CDS 3515. .5095
/note="GeneID exon scores (in order of location ranges): 144.31 - GSCJ_ID dd_02253"
/codon_start=1
/product="similar to Arabidopsis thaliana (Mouse-ear cress). Hypothetical 59.5 kDa protein"
/protein_id="AAO52156.1"
/db_xref="GI:28829640"
/translation="MNNQVKPTLSNKAEIINKWLIDNGVRINNKLIKIVYLGKENNFE
QTENTTATTTTSERINDSIVSGLGVISLKELKVDDIVAKIPKSIILSIHTSSISNILE
KYKIENNIGTSIALIHEASLGEKSKWYGYISSLPRKVDVPILWDSESRKLLKGTAIED
VLNDDDILINQVYADVIESILSKNHPEIFGDKELYSIENFKIANSIISSRAFCVDSYH
GDSLVPLADIFNHQTAREHVHIESNGDVCNKCGSIKTCKHRKVVTQHHTVNSTKGKRT
HKVAGIPSSKKHIHKGNCCSTTTTKTNEEDKDTIIEEDDEHLYIKVVKGVEANKEVYN
TYGDHDNAILLSKYGFLEMDNPCDRLSIDKQLVDKELELIGAENQITRMDLFNRISFY
AEMFDIDSRNNHAIENDGRLDDALVCSVGIALAPQSIFQSWKSMSEHKREKYFEQLEA
EDIVKQSQLVRNTLLSILNKLFSNYNNSNETTATTTIEQDQEKLSTLKNQREIIAQSL
KICEKTLINKSINYYKNL"
CDS complement(5235. .7895)
/note="GeneID exon scores (in order of location ranges): 270.65 - GSCJ_ID dd_03185"
/codon_start=1
/product="similar to essential for initiation of DNA replication; homolog of S. pombe CDC21; Cdc54p [Saccharomyces cerevisiae]"
/protein_id="AAO52157.1"
/db_xref="GI:28829641"
/translation="MDPKETKTSKISFLNTTKKNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNDDNNNNGENNIDIEPDDIDLLNKNDIDNIVDGIDNNNKSGSNDMKGLVF
WGTTVKVEDVKERFRKFIFNFPTPKERKRFNINDINNNNNNNNNNNSEEMNDDDDYYY
DEEQDKLIKKPEKVQYLYRDLLKVLNETKKCHLNINMKYLVQFDMELYLQWVSFPNEM
IPLLDEEINLIYRELFPITPRNGDDEDDDDDDEDEEDYRIELHPFNLLRKTPMRDLN
PSDIDKIISISGLIIRSSSIIPEIKQAFFMCAVCEATFHANVEKGKIQEPSECSNCKS
KQSLSIIHNRCLFGDKQYIKLQETPDAIPEGETPHTVALFAYGDLIDIAKPGDRVELT
GVFKASPMRAGSNRSLRSIYKTYIDILHIKRTDKGKYDDDDDHDDNTGGGGTGTGKE
TNENLDFEDLDEFDLSEEKEREIIELSKKPDIYDIVTKSIAPNIWELEDIKKGILCQL
FGGSKKSYQDYGGKFRGDINILLCGDPGTSKSQLLSYVHKIAPRGIYTSGKGSSAVGL
TAYITKDPDTKETVLESGALVLSDKGVCCIDEFDKMNDQTRSILHEVMEQQTVSIAKA
GIICTLNARTSILASANPSGSRYMPKLSVVENIQLPPTLLSRFDLIYLVLDKANERSD
RQLARHLVSMYWDETPVSHFTIPKETLTNYIQYARKHINPKLTDDSAKCLVQGYLEMR
SMGSSKKTISATPRQLESLIRIAEAHARIRFSEFVEPLDVEEAIRLIKVALQQAAIDP
ENGTIDMDLITTGRSASSREAITRLKSHIKQKIGKKHLTLDQLLKLLTIQNQVQTIQQ
IEEIKEALRQLQDEEIIQSSGGIDPTYGVV"
CDS join(8672. .9027,9184. .9463)
/note="GeneID exon scores (in order of location ranges): 37.11, 23.19 - GSCJ_ID dd_02251"
/codon_start=1
/product="similar to cyctochrome c1 heme lyase [Caenorhabditis elegans]"
/protein_id="AAO52158.1"
/db_xref="GI:28829642"
/translation="MGDDNTNNTSITKSSPTDAMGGGCPIVHDINPANHMYKPNQNPH
PDQTKPLSTERITSTIPRTEKDNWQYPSPQMFFNAMKKKQYEPKEEDMSVVISIHNTV
NEKCWEDVLQWENDYKDVCPNPKLKKFKGKATDFSPKAKFLNTFLGYKLPFDRHDWIV
DRNGKEVRYVIDFYEGRINKDSGKSIGIYIDVRPAIDDFSSLKDRVLHFFK"
CDS complement(join(9650. .13297,13390. .13536))
/note="GeneID exon scores (in order of location ranges): 468.70, 11.48 - GSCJ_ID dd_02250"
/codon_start=1
/product="similar to C25A11.4b.p [Caenorhabditis elegans]"
/protein_id="AAO52159.1"
/db_xref="GI:28829643"
/translation="MNPFEMIQNQNIVIKNAIEVGCIIVNIGAQDLMNATPYLVLGVI
WQIIKQGLLSKVNQNANEILEILFEEDSVKEVDNTQNQSDEHSAEEILLRWVNYHLEK
EGCERRISNFSEDIQDSVVYSHLFHQLVPIEFENLVSTIHSESNLFSRAELITKACEK
IGVKCFLTPSDIALGHPKLNLALVASLFNSEAAIANLRKEREEKHRTDQERLERERLE
RERIERERLERERLEQERLERERLERERLEKQRQEMERIEQERIELERLERERLEQER
LEQERIERERLERERLEQERLEQERLERERLEQERLECERLERERLEQERLEQERLER
ERIELEELERIERERLEYEERERIERERIERERREELEEQERLEYERLEQERIEREEME
RIERERIEREEFERLEEERIERERREEALRLEAERLERELRERERITRQSELERIEYNK
EYKICIRIQKRMIDDTSLTDNMTKLIGDITELKRNIIGELRNTTHHQAELKKLTKSID
RVLENKNEGKKLKKQKKVIIKPVKGVPNGLDPNKMKNYQSLFYYLQTRPEVMGKLLFL
LDRNQMGKTFDNLLLSMFPYEISPREEDLFLKALTVGIEYQIKYSETLQDFLQSESVP
MKLLYQYFKKKGFKYLKEIITPLILKVLDQTDLNLNFDPYWIQKEIQIAKELKDGKKS
EKKTLDQTFEQLMANNDVVSIFNSRCLQLVELGNAFLERITKSIEQLPVQLRYLCRII
AEKSKYAFPTNNTNNHNGGGISSDESGSTTANITLNNNRSKQLPENREKYISYFLII
YRFIIPIISSPDLLNELVLLQHPYITKQSRFNLSQLSRLIQSVFSEKPWMNYFPTVYL
NKWIIQNSSSADKFLCQVAQVPDLDELIHYSEYNEIMLENNIKFCSVINISEIIWLHE
TILSNFYDLEDIGTLKEVSEKSIQNLAERAINSKLDKGKYVTAARQSTPAIMTSQKSQ
NWPLKKYLDEIDFPIQLPEQGDRSIQLSFINRLDQLNEQSIDLWKECKLWMIDILSKS
QPPKKNANTIEIIREAIDYFKTGSNSSENLSIINKSNELISRLRTFDRDQCTDGLDGG
GPDNFNLFLLDVEKELATILRKNKVMRSELTRLHSLLKSIRIDAAPLEHLTKFQNEQL
EKAIVKKQLKTTQLTPLPKDKSTIKPHKFILLDLVKKGMILRCELPESSFSKTKIYIK
SLGPGLFEIEAKFGPITKSLQLELDDLLEKSKNHIDEFPLEGITLDVTLTIHLLMKLF
SL"
CDS complement(join(13685. .13812,14033. .14207))
/note="GeneID exon scores (in order of location ranges): 6.21, 14.57 - GSCJ_ID dd_02249"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAO52160.1"
/db_xref="GI:28829644"
/translation="MTYTDSRYNTLNRGSKKMDPPFITQEVLASAKISKKDNILINAN
AEDGSVHTFSEEEKVAYSKFITERLIDQEDVLSNYVPIDPSTDQLFVNCTDGVLLK"
CDS 16650. .17525
/note="GeneID exon scores (in order of location ranges): 78.29 - GSCJ_ID dd_00829"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAO52161.1"
/db_xref="GI:28829645"
/translation="METKDKKKDMELLGDRMKSYEDDMKIQIEKNKPFIIRLDGHSFS
KFTKNFNKPHDIRIHNAMIETSTVLLKTFMPTCIYTFSDEITMCFPSIDESTLEEGKE
IPNLAYSGKVQKLISLSSGLASTVFFKSITNAQYDKDTELNLIKLLETCTPHFDARIF
TLPSNQEIVNNLIWRSLVDCKRNSVANLGFAHFTPKQMLGLNNTEVKKKLLEEKSIIY
ENEPAWYRFGTYLKKEYYTLTTVSPNEPEKQITAIRSKVNCFSFDITKLSNSLNFISF
KTLPEGFNLELIESK"
CDS complement(join(17735. .18959,19168. .19640,19750. .20190, 20286. .20531))
/note="GeneID exon scores (in order of location ranges): 130.29, 48.89, 51.03, 20.75 - GSCJ_ID dd_00831"
/codon_start=1
/product="similar to Dictyostelium discoideum (Slime mold). ALG-2 interacting protein X"
/protein_id="AAO52162.1"
/db_xref="GI:28829646"
/translation="MLSIERKRTEKVDFSKPLTKYIKEQFSKAESDQHETQIATLNGL
REDVRNLQERTETSKEMVWKYYSILSSLELRFPISENNVRISFPWTDSYRQRKSTLYS
IYFERASVLFNYGSIVSQIASSTNRSNIEGVKKACNQFQLAAGVFNKLREYASLHPEC
STSADFSSESLQALVTIMLAQAQECIYEKASMDNLSDSILSKLAAQVAEYYDTFNQLL
NSNSLKSIVDRNWNITATVKSYLYKAISLYCHAKGLEQVSQFGEQVSRLMIAVDNINQ

```
                SKVNLAKTAPIELKEIVERYVISITRYCESAKKDNDTIYHDTIPPAHKLTPIEKKPLA
                KALPLPEINFVDPFNSLVPFSVKEDSAYYNDQKETLLRKELDNIEFHNQSAKASLLSM

Alignment Scores:
Pred. No.:              2.75            Length:         153751
Score:                  84.00           Matches:        12
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-858-332G-15 (1-12) x AC116551 (1-153751)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      73027 CACAACCACAACCACAACCACAACCACAACCACAAC 72992

RESULT 56
AY884207/c
LOCUS       AY884207              155983 bp    DNA     linear   PRI 23-JAN-2005
DEFINITION  Homo sapiens catenin (cadherin-associated protein), alpha 1, 102kDa
            (CTNNA1) gene, complete cds.
ACCESSION   AY884207
VERSION     AY884207.1  GI:57864639
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 155983)
  AUTHORS   Livingston,R.J., Rieder,M.J., Chung,M.-W., Downing,T.K.,
            Olson,A.N., Nguyen,C.P., Gildersleeve,H., Cassidy,C.M.,
            Johnson,E.J., Swanson,J.E., McFarland,I., Yool,B., Park,C. and
            Nickerson,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JAN-2005) Genome Sciences, University of Washington,
            1705 NE Pacific, Seattle, WA 98195, USA
COMMENT     To cite this work please use:  NIEHS-SNPs, Environmental Genome
            Project, NIEHS ES15478, Department of Genome Sciences, Seattle, WA
            (URL: http://egp.gs.washington.edu).
FEATURES             Location/Qualifiers
     source          1. .155983
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     variation       145. .147
                     /frequency="0.74"
                     /replace=""
     variation       347
                     /frequency="0.04"
                     /replace="g"
     variation       368
                     /frequency="0.01"
                     /replace="t"
     variation       431
                     /frequency="0.01"
                     /replace="a"
     repeat_region   464. .770
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     gene            <1117. .154041
                     /gene="CTNNA1"
     mRNA            join(<1117. .1224,2372. .2567,29235. .29401,31380. .31499,
                     43727. .43996,46713. .46916,105412. .105492,106690. .106842,
                     123550. .123642,136943. .137099,143712. .143912,
                     144458. .144609,148448. .148558,149675. .149856,
                     150032. .150137,151785. .151919,153009. .154041)
                     /gene="CTNNA1"
                     /product="catenin (cadherin-associated protein), alpha 1,
                     102kDa"
     CDS             join(1120. .1224,2372. .2567,29235. .29401,31380. .31499,
                     43727. .43996,46713. .46916,105412. .105492,106690. .106842,
                     123550. .123642,136943. .137099,143712. .143912,
                     144458. .144609,148448. .148558,149675. .149856,
                     150032. .150137,151785. .151919,153009. .153296)
                     /gene="CTNNA1"
                     /codon_start=1
                     /product="catenin (cadherin-associated protein), alpha 1,
                     102kDa"
                     /protein_id="AAW56940.1"
                     /db_xref="GI:57864640"
                     /translation="MTAVHAGNINFKWDPKSLEIRTLAVERLLEPLVTQVTTLVNTNS
                     KGPSNKKRGRSKKAHVLAASVEQATENFLEKGDKIAKESQFLKEELVAAVEDVRKQGD
                     LMKAAAGEFADDPCSSVKRGNMVRAARALLSAVTRLLILADMADVYKLLVQLKVVEDG
                     ILKLRNAGNEQDLGIQYKALKPEVDKLNIMAAKRQQELKDVGHRDQMAAARGILQKNV
                     PILYTASQACLQHPDVAAYKANRDLIYKQLQQAVTGISNAAQATASDDASQHQGGGGG
                     ELAYALNNFDKQIIVDPLSFSEERFRPSLEERLESIISGAALMADSSCTRDDRRERIV
                     AECNAVRQALQDLLSEYMGNAGRKERSDALNSAIDKMTKKTRDLRRQLRKAVMDHVSD
                     SFLETNVPLLVLIEAAKNGNEKEVKEYAQVFREHANKLIEVANLACSISNNEEGVKLV
                     RMSASQLEALCPQVINAALALAAKPQSKLAQENMDLFKEQWEKQVRVLTDAVDDITSI
                     DDFLAVSENHILEDVNKCVIALQEKDVDGLDRTAGAIRGRAARVIHVVTSEMDNYEPG
                     VYTEKVLEATKLLSNTVMPRFTEQVEAAVEALSSDPAQPMDENEFIDASRLVYDGIRD
                     IRKAVLMIRTPEELDDSDFETEDFDVRSRTSVQTEDDQLIAGQSARAIMAQLPQEQKA
                     KIAEQVASFQEEKSKLDAEVSKWDDSGNDIIVLAKQMCMIMMEMTDFTRGKGPLKNTS
                     DVISAAKKIAEAGSRMDKLGRTIADHCPDSACKQDLLAYLQRIALYCHQLNICSKVKA
                     EVQNLGGELVVSGVDSAMSLIQAAKNLMNAVVQTVKASYVASTKYQKSQGMASLNLPA
                     VSWKMKAPEKKPLVKREKQDETQTKIKRASQKKHVNPVQALSEFKAMDSI"
     variation       1361
                     /gene="CTNNA1"
                     /frequency="0.01"
                     /replace="g"
     variation       1438
                     /gene="CTNNA1"
                     /frequency="0.01"
                     /replace="t"
     variation       1444
                     /gene="CTNNA1"
                     /frequency="0.23"
                     /replace="a"
     variation       1841
                     /gene="CTNNA1"
                     /frequency="0.04"
                     /replace="t"
     variation       1844
                     /gene="CTNNA1"
                     /frequency="0.08"
                     /replace="t"
     variation       2906
                     /gene="CTNNA1"
                     /frequency="0.26"
                     /replace="a"
     repeat_region   3269. .3398
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     misc_feature    3504. .11400
                     /gene="CTNNA1"
                     /note="Region not scanned for variation"
     repeat_region   3785. .4230
                     /rpt_family="MaLR"
                     /rpt_type=dispersed
     repeat_region   4260. .4555
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   4749. .4934
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   4986. .5088
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   5637. .6101
                     /rpt_family="MaLR"
                     /rpt_type=dispersed
     repeat_region   6344. .6625
                     /rpt_family="Alu"
                     /rpt_type=dispersed
```

```
repeat_region   6626. .6749
                /rpt_family="MER1_type"
                /rpt_type=dispersed
repeat_region   7394. .7484
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   7623. .7786
                /rpt_family="L1"
                /rpt_type=dispersed
repeat_region   8649. .8950
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   9034. .9205
                /rpt_family="CR1"
                /rpt_type=dispersed
repeat_region   10532. .10837
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   11048. .11148
                /rpt_family="L1"
                /rpt_type=dispersed
variation       11585
                /gene="CTNNA1"
                /frequency="0.08"
                /replace="t"
variation       11907
                /gene="CTNNA1"
                /frequency="0.01"
                /replace="g"
variation       12040
                /gene="CTNNA1"
                /frequency="0.01"
                /replace="c"
variation       12051
                /gene="CTNNA1"
                /frequency="0.07"
                /replace="t"
misc_feature    12205. .12787
                /gene="CTNNA1"
                /note="Region not scanned for variation"
variation       13053
                /gene="CTNNA1"
                /frequency="0.01"
                /replace="c"
repeat_region   13208. .13511
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       13259
                /gene="CTNNA1"
                /frequency="0.02"
                /replace="t"
variation       13298
                /gene="CTNNA1"
                /frequency="0.01"
                /replace="t"
variation       13557
                /gene="CTNNA1"
                /frequency="0.07"
                /replace="c"
misc_feature    13683. .22861
                /gene="CTNNA1"
                /note="Region not scanned for variation"
repeat_region   14076. .14369
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   14375. .14676
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   15050. .15302
                /rpt_family="L1"
                /rpt_type=dispersed
repeat_region   15470. .15754
                /rpt_family="Alu"
```

```
Alignment Scores:
Pred. No.:              2.79          Length:         155983
Score:                  84.00         Matches:        12
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     8             Gaps:           0

US-09-858-332G-15 (1-12) x AY884207 (1-155983)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     113879 CATAATCATAATCATAATCATAATCATAATCATAAT 113844

RESULT 57
AC142155/c
LOCUS       AC142155              157087 bp    DNA     linear   HTG 24-MAR-2003
DEFINITION  Rattus norvegicus clone CH230-469L20, WORKING DRAFT SEQUENCE, 41
            unordered pieces.
ACCESSION   AC142155
VERSION     AC142155.2  GI:29165526
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 157087)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., Mcleod,M., Mcneill,T., Meenen,E., Milosavljevic,A.,
            Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K.,
            Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D.,
            Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G.,
            Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H.,
            Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A.,
            Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E.,
            Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y.,
            Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A.,
            Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S.,
            Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A.,
            Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E.,
            Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A.,
            Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S.,
            Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D.,
            Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J.,
            Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R.,
```

```
            Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S.,
            Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X.,
            Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R.,
            Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 157087)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAR-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 157087)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAR-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Mar 24, 2003 this sequence version replaced gi:29164535.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: KELQ
                Center clone name: CH230-469L20
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 135022 bases at least Q40
                Consensus quality: 140150 bases at least Q30
                Consensus quality: 144654 bases at least Q20
                Estimated insert size: 142503; sum-of-contigs estimation
                Quality coverage: 3x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 41 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1414: contig of 1414 bp in length
            *     1415     1514: gap of unknown length
            *     1515     2617: contig of 1103 bp in length
            *     2618     2717: gap of unknown length
            *     2718     4054: contig of 1337 bp in length
            *     4055     4154: gap of unknown length
            *     4155     5358: contig of 1204 bp in length
            *     5359     5458: gap of unknown length
            *     5459     6542: contig of 1084 bp in length
            *     6543     6642: gap of unknown length
            *     6643     7677: contig of 1035 bp in length
            *     7678     7777: gap of unknown length
            *     7778     8889: contig of 1112 bp in length
            *     8890     8989: gap of unknown length
            *     8990    10158: contig of 1169 bp in length
            *    10159    10258: gap of unknown length
            *    10259    12436: contig of 2178 bp in length
            *    12437    12536: gap of unknown length
            *    12537    13901: contig of 1365 bp in length
            *    13902    14001: gap of unknown length
            *    14002    15160: contig of 1159 bp in length
            *    15161    15260: gap of unknown length
            *    15261    16741: contig of 1481 bp in length
            *    16742    16841: gap of unknown length
            *    16842    18644: contig of 1803 bp in length
            *    18645    18744: gap of unknown length
            *    18745    20346: contig of 1602 bp in length
            *    20347    20446: gap of unknown length
            *    20447    22101: contig of 1655 bp in length
            *    22102    22201: gap of unknown length
            *    22202    23992: contig of 1791 bp in length
            *    23993    24092: gap of unknown length
            *    24093    26412: contig of 2320 bp in length
            *    26413    26512: gap of unknown length
            *    26513    27996: contig of 1484 bp in length
            *    27997    28096: gap of unknown length
            *    28097    29500: contig of 1404 bp in length
            *    29501    29600: gap of unknown length
            *    29601    31502: contig of 1902 bp in length
            *    31503    31602: gap of unknown length
            *    31603    34175: contig of 2573 bp in length
            *    34176    34275: gap of unknown length
            *    34276    36382: contig of 2107 bp in length
            *    36383    36482: gap of unknown length
            *    36483    38035: contig of 1553 bp in length
            *    38036    38135: gap of unknown length
            *    38136    40858: contig of 2723 bp in length
            *    40859    40958: gap of unknown length
            *    40959    45229: contig of 4271 bp in length
            *    45230    45329: gap of unknown length
            *    45330    49128: contig of 3799 bp in length
            *    49129    49228: gap of unknown length
            *    49229    54153: contig of 4925 bp in length
            *    54154    54253: gap of unknown length
            *    54254    60701: contig of 6448 bp in length
            *    60702    60801: gap of unknown length
            *    60802    65789: contig of 4988 bp in length
            *    65790    65889: gap of unknown length
            *    65890    70765: contig of 4876 bp in length
            *    70766    70865: gap of unknown length
            *    70866    75645: contig of 4780 bp in length
            *    75646    75745: gap of unknown length
            *    75746    81392: contig of 5647 bp in length
            *    81393    81492: gap of unknown length
            *    81493    89260: contig of 7768 bp in length
            *    89261    89360: gap of unknown length
            *    89361    94708: contig of 5348 bp in length
            *    94709    94808: gap of unknown length
            *    94809   103359: contig of 8551 bp in length
            *   103360   103459: gap of unknown length
            *   103460   111558: contig of 8099 bp in length
            *   111559   111658: gap of unknown length
            *   111659   118434: contig of 6776 bp in length
            *   118435   118534: gap of unknown length
            *   118535   127813: contig of 9279 bp in length
            *   127814   127913: gap of unknown length
            *   127914   137835: contig of 9922 bp in length
            *   137836   137935: gap of unknown length
            *   137936   147581: contig of 9646 bp in length
            *   147582   147681: gap of unknown length
            *   147682   157087: contig of 9406 bp in length.
FEATURES             Location/Qualifiers
     source          1. .157087
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-469L20"
     gap             1415. .1514
                     /estimated_length=unknown
     gap             2618. .2717
                     /estimated_length=unknown
     gap             4055. .4154
                     /estimated_length=unknown
     gap             5359. .5458
                     /estimated_length=unknown
     gap             6543. .6642
                     /estimated_length=unknown

Alignment Scores:
```

```
Pred. No.:              2.81          Length:        157087
Score:                  84.00         Matches:       12
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     14            Gaps:          0

US-09-858-332G-15 (1-12) x AC142155 (1-157087)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      20622 CACAACCACAACCACAACCACAACCACAACCACAAC 20587

RESULT 58
AL355315
LOCUS       AL355315              157193 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP11-548K23 on chromosome 10 Contains
            the ANKRD2 gene for ankyrin repeat domain 2 (stretch responsive
            muscle), six novel genes, the gene for phosphatidylinositol
            4-kinase type II (PI4KII) and four CpG islands, complete sequence.
ACCESSION   AL355315
VERSION     AL355315.18  GI:16972857
KEYWORDS    HTG; ANKRD2; ankyrin; CpG island; PI4KII.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 157193)
  AUTHORS   Chapman,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 16, 2001 this sequence version replaced gi:16580188.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 10, constructed by the Sanger Centre Chromosome 10
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr10
            RP11-548K23 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1. .157193
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-548K23"
                     /clone_lib="RPCI-11.2"
     misc_feature    1
                     /note="Clone_left_end: RP11-548K23"
     gene            join(AL359388.32:85573. .85949,6575. .6676,7015. .7173,
                     7273. .7380,9630. .9728,10091. .10189,11090. .11188,
                     11379. .11477,12351. .12659)
                     /gene="ANKRD2"
                     /locus_tag="RP11-548K23.2-001"
     gene            join(AL359388.32:85853. .85949,6575. .6676,7015. .7173,
                     7273. .7380,9630. .9728,10091. .10189,11379. .11477,
                     12351. .12659)
                     /gene="ANKRD2"
                     /locus_tag="RP11-548K23.2-002"
     mRNA            join(AL359388.32:85573. .85949,6575. .6676,7015. .7173,
                     7273. .7380,9630. .9728,10091. .10189,11090. .11188,
                     11379. .11477,12351. .12659)
                     /gene="ANKRD2"
                     /locus_tag="RP11-548K23.2-001"
                     /product="ankyrin repeat domain 2 (stretch responsive
                     muscle)"
                     /note="match: ESTs: AA346025.1 BF209399.1 BF790486.1
                     match: cDNAs: AB058599.1 AJ304805.1 AK056990.1"
     mRNA            join(AL359388.32:85853. .85949,6575. .6676,7015. .7173,
                     7273. .7380,9630. .9728,10091. .10189,11379. .11477,
                     12351. .12659)
                     /gene="ANKRD2"
                     /locus_tag="RP11-548K23.2-002"
                     /product="ankyrin repeat domain 2 (stretch responsive
                     muscle)"
                     /note="match: cDNAs: BC020817.1"
     CDS             join(AL359388.32:85863. .85949,6575. .6676,7015. .7173,
                     7273. .7380,9630. .9728,10091. .10189,11090. .11188,
                     11379. .11477,12351. .12500)
                     /gene="ANKRD2"
                     /locus_tag="RP11-548K23.2-001"
                     /standard_name="OTTHUMP00000020219"
                     /note="match: proteins: Q96MG0"
                     /codon_start=1
                     /product="ankyrin repeat domain 2 (stretch responsive
                     muscle)"
                     /protein_id="CAI15463.1"
                     /db_xref="GI:55958624"
                     /translation="MDGTMEDSEAVQRATALIEQRLAQEEENEKLRGDARQKLPMDLL
                     VLEDEKHHGAQSAALQKVKGQERVRKTSLDLRREIIDVGGIQNLIELRKKRKQKKRDA
                     LAASHEPPPEPEEITGPVDEETFLKAAVEGKMKVIEKFLADGGSADTCDQFRRTALHR
                     ASLEGHMEILEKLLDNGATVDFQDRLDCTAMHWACRGGHLEVVKLLQSHGADTNVRDK
                     LLSTPLHVAVRTGQVEIVEHFLSLGLEINARDREGDTALHDAVRLNRYKIIKLLLLHG
                     ADMMTKNLAGKTPTDLVQLWQADTRHALEHPEPGAEHNGLEGPNDSGRETPQPVPAQ"
     CDS             join(AL359388.32:85863. .85949,6575. .6676,7015. .7173,
                     7273. .7380,9630. .9728,10091. .10189,11379. .11477,
                     12351. .12500)
                     /gene="ANKRD2"
                     /locus_tag="RP11-548K23.2-002"
                     /standard_name="OTTHUMP00000020220"
                     /codon_start=1
                     /product="ankyrin repeat domain 2 (stretch responsive
                     muscle)"
                     /protein_id="CAI15462.1"
                     /db_xref="GI:55958623"
                     /db_xref="InterPro:IPR002110"
                     /db_xref="UniProt/TrEMBL:Q5T457"
                     /translation="MDGTMEDSEAVQRATALIEQRLAQEEENEKLRGDARQKLPMDLL
                     VLEDEKHHGAQSAALQKVKGQERVRKTSLDLRREIIDVGGIQNLIELRKKRKQKKRDA
                     LAASHEPPPEPEEITGPVDEETFLKAAVEGKMKVIEKFLADGGSADTCDQFRRTALHR
                     ASLEGHMEILEKLLDNGATVDFQDRLDCTAMHWACRGGHLEVVKLLQSHGADTNVRDK
                     EGDTALHDAVRLNRYKIIKLLLLHGADMMTKNLAGKTPTDLVQLWQADTRHALEHPEP
                     GAEHNGLEGPNDSGRETPQPVPAQ"
     polyA_signal    12637. .12642
                     /gene="ANKRD2"
                     /locus_tag="RP11-548K23.2-001"
     polyA_site      12659
                     /gene="ANKRD2"
                     /locus_tag="RP11-548K23.2-001"
     gene            join(13098. .13689,27550. .27678,27864. .29723)
                     /gene="C10orf65"
                     /locus_tag="RP11-548K23.9-002"
     mRNA            join(13098. .13689,27550. .27678,27864. .29723)
                     /gene="C10orf65"
```

/locus_tag="RP11-548K23.9-002"
/product="chromosome 10 open reading frame 65"
/note="match: cDNAs: AK094791.1"

gene join(13118. .13689,27550. .27678,27864. .27991,28455. .28589, 28842. .28938,30632. .30765,40285. .41577)
/gene="C10orf65"
/locus_tag="RP11-548K23.9-005"

gene join(13118. .13689,30632. .30765,40285. .41577)
/gene="C10orf65"
/locus_tag="RP11-548K23.9-001"

mRNA join(13118. .13689,27550. .27678,27864. .27991,28455. .28589, 28842. .28938,30632. .30765,40285. .41577)
/gene="C10orf65"
/locus_tag="RP11-548K23.9-005"
/product="chromosome 10 open reading frame 65"

mRNA join(13118. .13689,30632. .30765,40285. .41577)
/gene="C10orf65"
/locus_tag="RP11-548K23.9-001"
/product="chromosome 10 open reading frame 65"
/note="match: ESTs: BI547728.1
match: cDNAs: BC011916.1"

gene join(13459. .13689,27550. .27678,27864. .27991,28455. .28589, 28842. .28938,30632. .30765,40285. .40506)
/gene="C10orf65"
/locus_tag="RP11-548K23.9-003"

mRNA join(13459. .13689,27550. .27678,27864. .27991,28455. .28589, 28842. .28938,30632. .30765,40285. .40506)
/gene="C10orf65"
/locus_tag="RP11-548K23.9-003"
/product="chromosome 10 open reading frame 65"
/note="match: ESTs: AA065563.1 AI409706.1 BI549804.1 BI763832.1 BI764563.1 BI765630.1
match: cDNAs: AK002457.1 AK010857.1 BC016430.1"

CDS join(13479. .13689,27550. .27678,27864. .27991,28455. .28589, 28842. .28938,30632. .30765,40285. .40434)
/gene="C10orf65"
/locus_tag="RP11-548K23.9-003"
/standard_name="OTTHUMP00000020217"
/note="match: proteins: Q91W74 Q9CY60 Q9DCU9"
/codon_start=1
/product="chromosome 10 open reading frame 65"
/protein_id="CAI15454.1"
/db_xref="GI:55958615"
/db_xref="Genew:25155"
/db_xref="InterPro:IPR002220"
/db_xref="UniProt/TrEMBL:Q86XE5"
/translation="MLGPQVWSSVRQGLSRSLSRNVGVWASGEGKKVDIAGIYPPVTT
PFTATAEVDYGKLEENLHKLGTFPFRGFVVQGSNGEFPFLTSSERLEVVSRVRQAMPK
NRLLLAGSGCESTQATVEMTVSMAQVGADAAMVVTPCYYRGRMSSAALIHHYTKVADL
SPIPVVLYSVPANTGLDLPVDAVVTLSQHPNIVGMKDSGGDVTRIGLIVHKTRKQDFQ
VLAGSAGFLMASYALGAVGGVCALANVLGAQVCQLERLCCTGQWEDAQKLQHRLIEPN
AAVTRRFGIPGLKKIMDWFGYYGGPCRAPLQELSPAEEEALRMDFTSNGWL"

CDS join(13479. .13689,30632. .30765,40285. .40434)
/gene="C10orf65"
/locus_tag="RP11-548K23.9-001"
/standard_name="OTTHUMP00000020215"
/note="match: proteins: Q96EV5"
/codon_start=1
/product="chromosome 10 open reading frame 65"
/protein_id="CAI15455.1"
/db_xref="GI:55958616"
/db_xref="GOA:Q96EV5"
/db_xref="InterPro:IPR002220"
/db_xref="UniProt/TrEMBL:Q96EV5"
/translation="MLGPQVWSSVRQGLSRSLSRNVGVWASGEGKKVDIAGIYPPVTT
PFTATAEVDYGKLEENLHKLGTFPFRGAVGGVCALANVLGAQVCQLERLCCTGQWEDA
QKLQHRLIEPNAAVTRRFGIPGLKKIMDWFGYYGGPCRAPLQELSPAEEEALRMDFTS
NGWL"

CDS join(13479. .13689,27550. .27678,27864. .28060)

Alignment Scores:

Pred. No.:	2.81	Length:	157193
Score:	84.00	Matches:	12
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	8	Gaps:	0

US-09-858-332G-15 (1-12) x AL355315 (1-157193)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      55173 CATAATCATAATCATAATCATAATCATAATCATAAT 55208
```

RESULT 59
AC117110
LOCUS AC117110 161701 bp DNA linear HTG 19-NOV-2002
DEFINITION Rattus norvegicus clone CH230-234D18, WORKING DRAFT SEQUENCE, 2 unordered pieces.
ACCESSION AC117110
VERSION AC117110.4 GI:25072972
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 161701)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,

```
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 161701)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 161701)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 19, 2002 this sequence version replaced gi:23194554.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GTNJ
                Center clone name: CH230-234D18
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 153621 bases at least Q40
                Consensus quality: 154381 bases at least Q30
                Consensus quality: 154930 bases at least Q20
                Estimated insert size: 156926; sum-of-contigs estimation
                Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   157709: contig of 157709 bp in length
            *   157710   157809: gap of unknown length
            *   157810   161701: contig of 3892 bp in length.
FEATURES             Location/Qualifiers
     source          1. .161701
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-234D18"
     misc_feature    complement(609. .1485)
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:RWBOX21TJ"
     misc_feature    complement(149221. .150145)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:
                     end_sequence:RWBOX21TV"
     misc_feature    156629. .157709
                     /note="wgs_end_extension
                     clone_end:Sp6"
     gap             157710. .157809
                     /estimated_length=unknown
     misc_feature    157810. .158835
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    160542. .161701
                     /note="wgs_end_extension
                     clone_end:Sp6"
ORIGIN

Alignment Scores:
Pred. No.:              2.89          Length:         161701
Score:                  84.00         Matches:        12
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     14            Gaps:           0

US-09-858-332G-15 (1-12) x AC117110 (1-161701)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      65903 CACAACCACAACCACAACCACAACCACAACCACAAC 65938

RESULT 60
AC148008/c
LOCUS       AC148008              161965 bp    DNA     linear   ROD 29-MAY-2004
DEFINITION  Mus musculus BAC clone RP23-253C14 from chromosome 18, complete
            sequence.
ACCESSION   AC148008
VERSION     AC148008.2  GI:47131331
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 161965)
  AUTHORS   Waligorski,J. and Haakenson,W.
  TITLE     The sequence of Mus musculus BAC clone RP23-253C14
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 161965)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JAN-2004) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 161965)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAY-2004) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 161965)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAY-2004) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On May 12, 2004 this sequence version replaced gi:41351729.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: M_BA0253C14
            --------------.

            NOTICE:

            This sequence was finished as follows unless otherwise noted:
```

all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu

SOURCE INFORMATION:
The RPCI-23 BAC Library has been constructed by Kazutoyo Osegawa and Minako Tateno in the laboratory of Pieter de Jong (http://www.chori.org) from female C57BL/6J mouse kidney and/or brain genomic DNA. The clone and detailed information can be obtained from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at http://www.chori.org

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone. This clone is overlapped by AC144794 and AC120797.

FEATURES Location/Qualifiers

source	1. .161965
	/organism="Mus musculus"
	/mol_type="genomic DNA"
	/db_xref="taxon:10090"
	/chromosome="18"
	/map="18"
	/clone="RP23-253C14"
	/clone_lib="RPCI-23"
repeat_region	1605. .1681
	/rpt_family="MaLR"
repeat_region	1934. .2148
	/rpt_family="MaLR"
repeat_region	2430. .2822
	/rpt_family="MaLR"
repeat_region	2903. .3049
	/rpt_family="B4"
repeat_region	4067. .4150
	/rpt_family="MaLR"
repeat_region	4226. .4355
	/rpt_family="ERVK"
repeat_region	4376. .4523
	/rpt_family="Alu"
repeat_region	4748. .4950
	/rpt_family="B2"
repeat_region	5139. .5350
	/rpt_family="MaLR"
repeat_region	6762. .6854
	/rpt_family="L1"
repeat_region	6915. .7458
	/rpt_family="ERVK"
repeat_region	7585. .7813
	/rpt_family="MaLR"
repeat_region	8150. .8292
	/rpt_family="ERVL"
repeat_region	9900. .10058
	/rpt_family="L1"
repeat_region	10059. .10436
	/rpt_family="RMER14"
repeat_region	10437. .10563
	/rpt_family="L1"
repeat_region	10654. .10739
	/rpt_family="B2"
repeat_region	10740. .10929
	/rpt_family="ERVK"
repeat_region	11337. .11506
	/rpt_family="B2"
repeat_region	12767. .12826
	/rpt_family="MaLR"
repeat_region	12870. .13089
	/rpt_family="ERVL"
repeat_region	13090. .13289
	/rpt_family="MaLR"
repeat_region	13300. .13444
	/rpt_family="Alu"
repeat_region	14035. .14083
	/rpt_family="ERV1"
repeat_region	15986. .16411
	/rpt_family="MaLR"
repeat_region	17388. .17450
	/rpt_family="ERV1"
repeat_region	17657. .17835
	/rpt_family="MaLR"
repeat_region	17836. .18235
	/rpt_family="MaLR"
repeat_region	18236. .18276
	/rpt_family="MaLR"
repeat_region	18564. .19053
	/rpt_family="L1"
repeat_region	20485. .20742
	/rpt_family="L1"
repeat_region	21352. .21835
	/rpt_family="MaLR"
repeat_region	21813. .22021
	/rpt_family="MaLR"
repeat_region	22022. .22221
	/rpt_family="MER2_type"
repeat_region	22373. .22515
	/rpt_family="MER2_type"
repeat_region	22723. .23139
	/rpt_family="RMER19B"
repeat_region	23132. .23789
	/rpt_family="RMER19B"
repeat_region	26200. .26479
	/rpt_family="MaLR"
repeat_region	26480. .27094
	/rpt_family="MaLR"
repeat_region	27193. .27638
	/rpt_family="MaLR"
repeat_region	27636. .27880
	/rpt_family="MaLR"
repeat_region	27943. .27995
	/rpt_family="ERV1"
repeat_region	28616. .28762
	/rpt_family="Alu"
repeat_region	29967. .30253
	/rpt_family="B4"
repeat_region	32879. .32988
	/rpt_family="MaLR"
repeat_region	33621. .35093
	/rpt_family="RMER1A"
repeat_region	35367. .35625
	/rpt_family="MaLR"
repeat_region	36503. .36910
	/rpt_family="L1"
repeat_region	36911. .37054
	/rpt_family="Alu"
repeat_region	37055. .37070
	/rpt_family="L1"
repeat_region	37061. .38066
	/rpt_family="L1"
repeat_region	38065. .39182
	/rpt_family="L1"
repeat_region	39173. .39733
	/rpt_family="L1"
repeat_region	39734. .39924
	/rpt_family="B2"
repeat_region	39925. .40025
	/rpt_family="L1"
repeat_region	43573. .43718
	/rpt_family="Alu"
repeat_region	44675. .45149

```
                     /rpt_family="ERV1"
     repeat_region   45199. .45518
                     /rpt_family="L1"
     repeat_region   46734. .46856
                     /rpt_family="Alu"
     repeat_region   47591. .47785
                     /rpt_family="B2"
     repeat_region   47804. .47997
                     /rpt_family="RMER17C"
     repeat_region   49007. .49181

Alignment Scores:
Pred. No.:              2.89          Length:         161965
Score:                  84.00         Matches:        12
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     9             Gaps:           0

US-09-858-332G-15 (1-12) x AC148008 (1-161965)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     141154 CACAACCACAACCACAACCACAACCACAACCACAAC 141119

RESULT 61
AC139383/c
LOCUS       AC139383              163087 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-232M19, *** SEQUENCING IN PROGRESS
            ***, 76 unordered pieces.
ACCESSION   AC139383
VERSION     AC139383.5  GI:30521079
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 163087)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., Mcleod,M., Mcneill,T., Meenen,E., Milosavljevic,A.,
            Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K.,
            Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D.,
            Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G.,
            Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H.,
            Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A.,
            Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E.,
            Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y.,
            Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A.,
            Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S.,
            Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A.,
            Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E.,
            Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A.,
            Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S.,
            Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D.,
            Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J.,
            Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R.,
            Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S.,
            Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X.,
            Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R.,
            Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163087)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-FEB-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 163087)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:30466920.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: KDIX
                Center clone name: CH230-232M19
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 125757 bases at least Q40
                Consensus quality: 133178 bases at least Q30
                Consensus quality: 138608 bases at least Q20
                Estimated insert size: 121885; sum-of-contigs estimation
                Quality coverage: 2x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 76 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1163: contig of 1163 bp in length
            *     1164     1263: gap of unknown length
            *     1264     2405: contig of 1142 bp in length
            *     2406     2505: gap of unknown length
            *     2506     3595: contig of 1090 bp in length
            *     3596     3695: gap of unknown length
            *     3696     4716: contig of 1021 bp in length
            *     4717     4816: gap of unknown length
            *     4817     6181: contig of 1365 bp in length
            *     6182     6281: gap of unknown length
            *     6282     7973: contig of 1692 bp in length
            *     7974     8073: gap of unknown length
            *     8074     9638: contig of 1565 bp in length
            *     9639     9738: gap of unknown length
            *     9739    11399: contig of 1661 bp in length
            *    11400    11499: gap of unknown length
```

```
*    11500    12507: contig of 1008 bp in length
*    12508    12607: gap of unknown length
*    12608    13877: contig of 1270 bp in length
*    13878    13977: gap of unknown length
*    13978    15121: contig of 1144 bp in length
*    15122    15221: gap of unknown length
*    15222    16241: contig of 1020 bp in length
*    16242    16341: gap of unknown length
*    16342    17364: contig of 1023 bp in length
*    17365    17464: gap of unknown length
*    17465    18591: contig of 1127 bp in length
*    18592    18691: gap of unknown length
*    18692    20093: contig of 1402 bp in length
*    20094    20193: gap of unknown length
*    20194    21253: contig of 1060 bp in length
*    21254    21353: gap of unknown length
*    21354    22737: contig of 1384 bp in length
*    22738    22837: gap of unknown length
*    22838    23886: contig of 1049 bp in length
*    23887    23986: gap of unknown length
*    23987    25395: contig of 1409 bp in length
*    25396    25495: gap of unknown length
*    25496    26906: contig of 1411 bp in length
*    26907    27006: gap of unknown length
*    27007    28457: contig of 1451 bp in length
*    28458    28557: gap of unknown length
*    28558    30511: contig of 1954 bp in length
*    30512    30611: gap of unknown length
*    30612    32038: contig of 1427 bp in length
*    32039    32138: gap of unknown length
*    32139    33355: contig of 1217 bp in length
*    33356    33455: gap of unknown length
*    33456    34544: contig of 1089 bp in length
*    34545    34644: gap of unknown length
*    34645    36300: contig of 1656 bp in length
*    36301    36400: gap of unknown length
*    36401    38424: contig of 2024 bp in length
*    38425    38524: gap of unknown length
*    38525    40057: contig of 1533 bp in length
*    40058    40157: gap of unknown length
*    40158    41398: contig of 1241 bp in length
*    41399    41498: gap of unknown length
*    41499    42538: contig of 1040 bp in length
*    42539    42638: gap of unknown length
*    42639    44986: contig of 2348 bp in length
*    44987    45086: gap of unknown length
*    45087    46574: contig of 1488 bp in length
*    46575    46674: gap of unknown length
*    46675    48143: contig of 1469 bp in length
*    48144    48243: gap of unknown length
*    48244    49755: contig of 1512 bp in length
*    49756    49855: gap of unknown length
*    49856    50905: contig of 1050 bp in length
*    50906    51005: gap of unknown length
*    51006    52535: contig of 1530 bp in length
*    52536    52635: gap of unknown length
*    52636    54233: contig of 1598 bp in length
*    54234    54333: gap of unknown length
*    54334    55448: contig of 1115 bp in length
*    55449    55548: gap of unknown length
*    55549    56713: contig of 1165 bp in length
*    56714    56813: gap of unknown length
*    56814    58783: contig of 1970 bp in length
*    58784    58883: gap of unknown length
*    58884    59967: contig of 1084 bp in length
*    59968    60067: gap of unknown length
*    60068    61165: contig of 1098 bp in length
*    61166    61265: gap of unknown length
*    61266    62383: contig of 1118 bp in length
*    62384    62483: gap of unknown length
*    62484    64426: contig of 1943 bp in length
*    64427    64526: gap of unknown length
*    64527    67010: contig of 2484 bp in length
*    67011    67110: gap of unknown length
*    67111    68249: contig of 1139 bp in length
*    68250    68349: gap of unknown length
*    68350    69652: contig of 1303 bp in length
*    69653    69752: gap of unknown length
*    69753    71827: contig of 2075 bp in length
*    71828    71927: gap of unknown length
*    71928    73967: contig of 2040 bp in length

Alignment Scores:
Pred. No.:                2.91          Length:          163087
Score:                    84.00         Matches:         12
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       14            Gaps:            0

US-09-858-332G-15 (1-12) x AC139383 (1-163087)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     150701 CACAACCACAACCACAACCACAACCACAACCACAAC 150666

RESULT 62
AC159252
LOCUS       AC159252               168456 bp    DNA     linear   ROD 29-JUN-2005
DEFINITION  Mus musculus BAC clone RP24-170P9 from chromosome 13, complete
            sequence.
ACCESSION   AC159252 AC139220
VERSION     AC159252.3  GI:68303781
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 168456)
  AUTHORS   Adeniyi,B., Cotton,M. and Levy,A.
  TITLE     The sequence of Mus musculus BAC clone RP24-170P9
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 168456)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-2005) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 168456)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAY-2005) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 168456)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-2005) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jun 29, 2005 this sequence version replaced gi:66793604.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: M_BB0170P09
            Drafting center: WIBR

            NOTICE:

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e. phred quality
            >=30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
```

one plasmid subclone, fosmid clone or direct clone walk sequence. Sequence from the Mouse Genome Sequencing Consortium whole genome shotgun may have been used to obtain the consensus sequence. The assembly was confirmed by restriction digest.
This finishing standard has slightly changed from the previous Human standard. Specifically, standards for regions of low sequence complexity (such as dinucleotide repeats and small unit tandem repeats) have been relaxed. These regions are very prevalent in the mouse genome, and the return on extended finishing efforts is minimal.
If a sequence meets the criteria of the above statement, it needs no comments or tags. If the criteria are not met, such as ambiguous bases, then the region is duly annotated.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu

SOURCE INFORMATION:
The BAC Library has been constructed by Pieter de Jong and coworkers (http://www.chori.org) from male C57BL/6J mouse spleen and/or brain genomic DNA. The clone and detailed information can be obtained from Pieter de Jong and coworkers at http://www.chori.org

This sequence is the entire insert of the clone.

```
FEATURES             Location/Qualifiers
     source          1. .168456
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="13"
                     /clone="RP24-170P9"
                     /clone_lib="RPCI-24"
     unsure          84508. .84814
                     /note="Unresolved simple sequence repeat."
     misc_feature    121486. .121500
                     /note="Sequence derived from PCR product of project DNA"
     unsure          156146. .156321
                     /note="Unresolved simple sequence repeat."
ORIGIN

Alignment Scores:
Pred. No.:              3.01          Length:        168456
Score:                  84.00         Matches:       12
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     9             Gaps:          0

US-09-858-332G-15 (1-12) x AC159252 (1-168456)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     161302 CACAACCACAACCACAACCACAACCACAACCACAAC 161337

RESULT 63
AP003356/c
LOCUS       AP003356              170533 bp    DNA     linear   PRI 28-JUN-2001
DEFINITION  Homo sapiens genomic DNA, chromosome 8q23, clone: KB1732A1.
ACCESSION   AP003356
VERSION     AP003356.2  GI:14572662
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Shimizu,N. and Asakawa,S.
  TITLE     Homo sapiens DNA chromosome 8 SEQUENCE
  JOURNAL   Published Only in DataBase (2001)
REFERENCE   2  (bases 1 to 170533)
  AUTHORS   Shimizu,N. and Asakawa,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-FEB-2001) Nobuyoshi Shimizu, Keio University, School
            of Medicine, Molecular Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo
            160-8582, Japan (E-mail:nshimizu@dmb.med.keio.ac.jp,
            Tel:81-3-3351-2370, Fax:81-3-3351-2370)
COMMENT     On Jun 27, 2001 this sequence version replaced gi:13359071.
FEATURES             Location/Qualifiers
     source          1. .170533
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8q23"
                     /clone="KB1732A1"
                     /cell_line="FLEB 14 - 14"
                     /clone_lib="Keio BAC library"
     repeat_region   88. .124
                     /evidence=not_experimental
                     /rpt_family="(CA)n"
     repeat_region   complement(185. .493)
                     /evidence=not_experimental
                     /rpt_family="AluSg"
     repeat_region   557. .843
                     /evidence=not_experimental
                     /rpt_family="MLT2H"
     repeat_region   925. .1076
                     /evidence=not_experimental
                     /rpt_family="MLT2H"
     repeat_region   1639. .1789
                     /evidence=not_experimental
                     /rpt_family="L2"
     repeat_region   2432. .2454
                     /evidence=not_experimental
                     /rpt_family="(TTTA)n"
     repeat_region   complement(2455. .2732)
                     /evidence=not_experimental
                     /rpt_family="AluJo"
     repeat_region   3010. .3115
                     /evidence=not_experimental
                     /rpt_family="MIR"
     repeat_region   3206. .3315
                     /evidence=not_experimental
                     /rpt_family="L2"
     repeat_region   3518. .3558
                     /evidence=not_experimental
                     /rpt_family="(CAAA)n"
     repeat_region   3863. .3945
                     /evidence=not_experimental
                     /rpt_family="L2"
     repeat_region   3995. .4224
                     /evidence=not_experimental
                     /rpt_family="MER20"
     repeat_region   4337. .4371
                     /evidence=not_experimental
                     /rpt_family="(CA)n"
     repeat_region   complement(4418. .4718)
                     /evidence=not_experimental
                     /rpt_family="AluSg"
     repeat_region   complement(4719. .5018)
                     /evidence=not_experimental
                     /rpt_family="AluJb"
     repeat_region   complement(5089. .5139)
                     /evidence=not_experimental
                     /rpt_family="L2"
     repeat_region   5377. .5505
                     /evidence=not_experimental
                     /rpt_family="L2"
     repeat_region   complement(6473. .6777)
                     /evidence=not_experimental
                     /rpt_family="AluSx"
     repeat_region   complement(7730. .8029)
```

```
                     /evidence=not_experimental
                     /rpt_family="AluSq"
     repeat_region   8637. .8762
                     /evidence=not_experimental
                     /rpt_family="FLAM_A"
     repeat_region   complement(9317. .9514)
                     /evidence=not_experimental
                     /rpt_family="MER63A"
     repeat_region   9594. .9905
                     /evidence=not_experimental
                     /rpt_family="AluY"
     repeat_region   9915. .10226
                     /evidence=not_experimental
                     /rpt_family="AluSx"
     repeat_region   complement(10429. .10716)
                     /evidence=not_experimental
                     /rpt_family="MLT1J"
     repeat_region   10890. .11118
                     /evidence=not_experimental
                     /rpt_family="MIR"
     repeat_region   11284. .11437
                     /evidence=not_experimental
                     /rpt_family="MIR"
     repeat_region   11576. .12029
                     /evidence=not_experimental
                     /rpt_family="LOR1b"
     repeat_region   12101. .12333
                     /evidence=not_experimental
                     /rpt_family="L2"
     repeat_region   13245. .13330
                     /evidence=not_experimental
                     /rpt_family="MIR"
     repeat_region   13414. .13455
                     /evidence=not_experimental
                     /rpt_family="MIR"
     repeat_region   13483. .13601
                     /evidence=not_experimental
                     /rpt_family="FLAM_A"
     repeat_region   complement(13641. .13915)
                     /evidence=not_experimental
                     /rpt_family="AluSx"
     repeat_region   complement(13942. .14081)
                     /evidence=not_experimental
                     /rpt_family="L2"
     repeat_region   14114. .14147
                     /evidence=not_experimental
                     /rpt_family="(TG)n"
     repeat_region   14529. .14682
                     /evidence=not_experimental
                     /rpt_family="MIR"
     repeat_region   15265. .15305
                     /evidence=not_experimental
                     /rpt_family="AT_rich"
     repeat_region   15606. .15627
                     /evidence=not_experimental
                     /rpt_family="AT_rich"
     repeat_region   16299. .16513
                     /evidence=not_experimental
                     /rpt_family="MER20"
     repeat_region   16554. .16845
                     /evidence=not_experimental
                     /rpt_family="AluSx"
     repeat_region   16879. .17181
                     /evidence=not_experimental
                     /rpt_family="AluYb8"
     repeat_region   17248. .17268
                     /evidence=not_experimental
                     /rpt_family="AT_rich"
     repeat_region   17408. .17538
                     /evidence=not_experimental
                     /rpt_family="MIR"
     repeat_region   17833. .18144
                     /evidence=not_experimental
                     /rpt_family="AluSq"
     repeat_region   18162. .18188
                     /evidence=not_experimental
                     /rpt_family="(TTTG)n"
     repeat_region   complement(19558. .19961)
                     /evidence=not_experimental
                     /rpt_family="MLT1A1"
     repeat_region   complement(19967. .20661)
                     /evidence=not_experimental
                     /rpt_family="L2"
     repeat_region   21023. .21060
                     /evidence=not_experimental
                     /rpt_family="(TG)n"
     repeat_region   21204. .21312
                     /evidence=not_experimental
                     /rpt_family="MIR"
     repeat_region   complement(21886. .22132)
                     /evidence=not_experimental
                     /rpt_family="MIR"
     repeat_region   complement(23127. .23412)
                     /evidence=not_experimental
                     /rpt_family="L1MC5"
     repeat_region   23422. .23483
                     /evidence=not_experimental
                     /rpt_family="FLAM_C"
     repeat_region   23484. .23560
                     /evidence=not_experimental
                     /rpt_family="LOR1a"
     repeat_region   complement(23561. .23865)
                     /evidence=not_experimental
                     /rpt_family="AluSg"
     repeat_region   23866. .24295
                     /evidence=not_experimental
                     /rpt_family="LOR1a"
     repeat_region   24296. .24352
                     /evidence=not_experimental
                     /rpt_family="FLAM_C"
     repeat_region   complement(24376. .24515)
                     /evidence=not_experimental
                     /rpt_family="L1MC5"

Alignment Scores:
Pred. No.:              3.04          Length:         170533
Score:                  84.00         Matches:        12
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     8             Gaps:           0

US-09-858-332G-15 (1-12) x AP003356 (1-170533)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      50732 CATAATCATAATCATAATCATAATCATAATCATAAT 50697

RESULT 64
AC068001/c
LOCUS       AC068001              176733 bp    DNA     linear   HTG 24-AUG-2002
DEFINITION  Homo sapiens chromosome 16 clone RP11-303C4 map 16, WORKING DRAFT
            SEQUENCE, 19 unordered pieces.
ACCESSION   AC068001
VERSION     AC068001.2  GI:8139051
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 176733)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 16, clone RP11-303C4
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 176733)
```

```
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
          Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
          Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
          Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
          Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
          Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
          Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
          Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
          Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
          Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
          Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
          McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
          Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
          Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
          O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
          Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
          Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
          Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
          Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
          Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
          Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE     Direct Submission
JOURNAL   Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3  (bases 1 to 176733)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
          Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
          Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
          Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
          Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
          Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
          Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
          Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
          Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
          Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
          Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
          McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
          Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
          Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
          O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
          Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
          Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
          Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
          Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
          Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
          Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE     Direct Submission
JOURNAL   Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Jun 1, 2000 this sequence version replaced gi:7656781.
          All repeats were identified using RepeatMasker:
          Smit, A.F.A. & Green, P. (1996-1997)
          http://ftp.genome.washington.edu/RM/RepeatMasker.html
          -------------- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
          -------------- Project Information
              Center project name: L8119
              Center clone name: 303_C_4
          -------------- Summary Statistics
              Sequencing vector: M13; M77815; 100% of reads
              Chemistry: Dye-terminator Big Dye; 100% of reads
              Assembly program: Phrap; version 0.960731
              Consensus quality: 165466 bases at least Q40
              Consensus quality: 171544 bases at least Q30
              Consensus quality: 173762 bases at least Q20
              Insert size: 182000; agarose-fp
              Insert size: 174933; sum-of-contigs
              Quality coverage: 4.3 in Q20 bases; agarose-fp
              Quality coverage: 4.5 in Q20 bases; sum-of-contigs
          --------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 19 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1     1180: contig of 1180 bp in length
          *     1181     1280: gap of 100 bp
          *     1281     2705: contig of 1425 bp in length
          *     2706     2805: gap of 100 bp
          *     2806     4387: contig of 1582 bp in length
          *     4388     4487: gap of 100 bp
          *     4488     6165: contig of 1678 bp in length
          *     6166     6265: gap of 100 bp
          *     6266    10875: contig of 4610 bp in length
          *    10876    10975: gap of 100 bp
          *    10976    17071: contig of 6096 bp in length
          *    17072    17171: gap of 100 bp
          *    17172    23490: contig of 6319 bp in length
          *    23491    23590: gap of 100 bp
          *    23591    29158: contig of 5568 bp in length
          *    29159    29258: gap of 100 bp
          *    29259    36785: contig of 7527 bp in length
          *    36786    36885: gap of 100 bp
          *    36886    44832: contig of 7947 bp in length
          *    44833    44932: gap of 100 bp
          *    44933    52974: contig of 8042 bp in length
          *    52975    53074: gap of 100 bp
          *    53075    62479: contig of 9405 bp in length
          *    62480    62579: gap of 100 bp
          *    62580    71273: contig of 8694 bp in length
          *    71274    71373: gap of 100 bp
          *    71374    79376: contig of 8003 bp in length
          *    79377    79476: gap of 100 bp
          *    79477    90116: contig of 10640 bp in length
          *    90117    90216: gap of 100 bp
          *    90217   105794: contig of 15578 bp in length
          *   105795   105894: gap of 100 bp
          *   105895   124221: contig of 18327 bp in length
          *   124222   124321: gap of 100 bp
          *   124322   146798: contig of 22477 bp in length
          *   146799   146898: gap of 100 bp
          *   146899   176733: contig of 29835 bp in length.
FEATURES             Location/Qualifiers
     source          1. .176733
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16"
                     /clone="RP11-303C4"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1180
                     /note="assembly_fragment"
     gap             1181. .1280
                     /estimated_length=100
     misc_feature    1281. .2705
                     /note="assembly_fragment"
     gap             2706. .2805
                     /estimated_length=100
     misc_feature    2806. .4387
                     /note="assembly_fragment"
     gap             4388. .4487
                     /estimated_length=100
     misc_feature    4488. .6165
                     /note="assembly_fragment"
     gap             6166. .6265
                     /estimated_length=100
     misc_feature    6266. .10875
                     /note="assembly_fragment"
     gap             10876. .10975
```

```
                     /estimated_length=100
     misc_feature    10976. .17071
                     /note="assembly_fragment"
     gap             17072. .17171
                     /estimated_length=100
     misc_feature    17172. .23490
                     /note="assembly_fragment"
     gap             23491. .23590
                     /estimated_length=100
     misc_feature    23591. .29158
                     /note="assembly_fragment"
     gap             29159. .29258
                     /estimated_length=100
     misc_feature    29259. .36785
                     /note="assembly_fragment"
     gap             36786. .36885
                     /estimated_length=100
     misc_feature    36886. .44832
                     /note="assembly_fragment"
     gap             44833. .44932
                     /estimated_length=100
     misc_feature    44933. .52974
                     /note="assembly_fragment"
     gap             52975. .53074
                     /estimated_length=100
     misc_feature    53075. .62479
                     /note="assembly_fragment"
     gap             62480. .62579
                     /estimated_length=100
     misc_feature    62580. .71273
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     gap             71274. .71373
                     /estimated_length=100
     misc_feature    71374. .79376
                     /note="assembly_fragment"
     gap             79377. .79476

Alignment Scores:
Pred. No.:              3.15          Length:         176733
Score:                  84.00         Matches:        12
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     14            Gaps:           0

US-09-858-332G-15 (1-12) x AC068001 (1-176733)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     143301 CATAATCATAATCATAATCATAATCATAATCATAAT 143266

RESULT 65
AC139578/c
LOCUS       AC139578              179415 bp    DNA     linear   ROD 10-JUL-2004
DEFINITION  Mus musculus BAC clone RP23-188P12 from chromosome 18, complete
            sequence.
ACCESSION   AC139578
VERSION     AC139578.4  GI:48675543
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 179415)
  AUTHORS   Kim,K., Kozlowicz,A., Haglund,K. and Bielicki,L.
  TITLE     The sequence of Mus musculus BAC clone RP23-188P12
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 179415)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-FEB-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 179415)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-AUG-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 179415)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-2004) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   5  (bases 1 to 179415)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2004) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jun 12, 2004 this sequence version replaced gi:33469236.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: M_BA0188P12
            --------------.

            NOTICE:

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-23 BAC Library has been constructed by Kazutoyo Osegawa
            and Minako Tateno in the laboratory of Pieter de Jong
            (http://www.chori.org) from female C57BL/6J mouse kidney and/or
            brain genomic DNA.  The clone and detailed information can be
            obtained from Research Genetics, Inc. (http://www.resgen.com) or
            Pieter de Jong and coworkers at http://www.chori.org

            NEIGHBORING SEQUENCE INFORMATION:
            This sequence is the entire insert of the clone.
FEATURES                Location/Qualifiers
     source             1. .179415
                        /organism="Mus musculus"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:10090"
                        /chromosome="18"
                        /map="18"
                        /clone="RP23-188P12"
                        /clone_lib="RPCI-23"
     repeat_region      1687. .1871
                        /rpt_family="CR1"
     repeat_region      2929. .3440
                        /rpt_family="ERVK"
     repeat_region      3432. .3626
                        /rpt_family="ERVK"
     repeat_region      4403. .4601
                        /rpt_family="MaLR"
     repeat_region      4610. .4884
                        /rpt_family="L1"
```

```
misc_feature    4932. .6098
                /note="CpG island (%GC=66.2, o/e=0.65, #CpGs=79)"
repeat_region   17342. .17772
                /rpt_family="L1"
repeat_region   17773. .17972
                /rpt_family="MaLR"
repeat_region   19555. .19738
                /rpt_family="B2"
repeat_region   20825. .20965
                /rpt_family="Alu"
repeat_region   21192. .21269
                /rpt_family="L1"
repeat_region   21417. .21736
                /rpt_family="ERV1"
repeat_region   21737. .22120
                /rpt_family="MaLR"
repeat_region   22121. .22285
                /rpt_family="ERV1"
repeat_region   22561. .22708
                /rpt_family="MaLR"
repeat_region   24235. .24415
                /rpt_family="B2"
repeat_region   24428. .24727
                /rpt_family="MaLR"
repeat_region   24729. .24802
                /rpt_family="L1"
repeat_region   25431. .25507
                /rpt_family="ID"
repeat_region   25514. .25711
                /rpt_family="MER1_type"
repeat_region   27264. .27464
                /rpt_family="MER1_type"
repeat_region   27747. .28051
                /rpt_family="MaLR"
repeat_region   28182. .28467
                /rpt_family="B4"
repeat_region   28593. .28971
                /rpt_family="L1"
repeat_region   30144. .30479
                /rpt_family="L1"
repeat_region   31424. .31643
                /rpt_family="L1"
repeat_region   31885. .32259
                /rpt_family="MaLR"
repeat_region   32590. .32663
                /rpt_family="BC1_MM"
repeat_region   32667. .32779
                /rpt_family="Alu"
repeat_region   33647. .33822
                /rpt_family="B2"
repeat_region   35418. .35579
                /rpt_family="B4"
repeat_region   35692. .36238
                /rpt_family="L1"
repeat_region   39587. .39741
                /rpt_family="B4"
repeat_region   39742. .39966
                /rpt_family="MER1_type"
repeat_region   39967. .40067
                /rpt_family="B4"
repeat_region   40035. .40094
                /rpt_family="B4"
repeat_region   40132. .40256
                /rpt_family="RMER1A"
repeat_region   40737. .41031
                /rpt_family="L1"
repeat_region   41053. .41227
                /rpt_family="B4"
repeat_region   42103. .42298
                /rpt_family="MER1_type"
unsure          42942. .42972
                /note="Sequence derived from one plasmid subclone."
repeat_region   43203. .43518
                /rpt_family="B4"
repeat_region   44343. .44554
                /rpt_family="B4"
repeat_region   45269. .45477
                /rpt_family="B2"
repeat_region   45786. .45922
                /rpt_family="Alu"
repeat_region   45927. .45995
                /rpt_family="B4"
repeat_region   46646. .46792
                /rpt_family="Alu"
repeat_region   46973. .47231
                /rpt_family="B4"
repeat_region   53665. .53772
                /rpt_family="B2"
repeat_region   54230. .54294
                /rpt_family="ERV1"
repeat_region   54300. .54388
                /rpt_family="Alu"
repeat_region   54302. .54413
                /rpt_family="B4"
repeat_region   54566. .54774
                /rpt_family="B4"
repeat_region   54936. .54991
                /rpt_family="MaLR"
repeat_region   61258. .61600
                /rpt_family="MaLR"
repeat_region   61649. .61795
                /rpt_family="MIR"
repeat_region   61954. .62184
                /rpt_family="MER1_type"
repeat_region   63519. .63592
                /rpt_family="B2"
repeat_region   63593. .63718
                /rpt_family="B4"
repeat_region   63729. .63837

Alignment Scores:
Pred. No.:              3.19          Length:         179415
Score:                  84.00         Matches:        12
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     9             Gaps:           0

US-09-858-332G-15 (1-12) x AC139578 (1-179415)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      82863 CACAACCACAACCACAACCACAACCACAACCACAAC 82828

RESULT 66
AC164547/c
LOCUS       AC164547              179452 bp    DNA     linear   HTG 30-JUL-2005
DEFINITION  Mus musculus chromosome 13 clone RP24-198D11, WORKING DRAFT
            SEQUENCE, 3 unordered pieces.
ACCESSION   AC164547
VERSION     AC164547.3  GI:71533382
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 179452)
  AUTHORS   Wilson,R.K.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 179452)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUN-2005) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
```

```
REFERENCE   3  (bases 1 to 179452)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUL-2005) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Jul 30, 2005 this sequence version replaced gi:68226608.

            ----------------- Genome Center ---------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu
            Contact: submissions@watson.wustl.edu
            -------------- Project Information ------------------
            Center project name: M_BB0198D11
            ----------------------------------------------------
            ------------- Summary Statistics  -------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 178552 bases at least Q40
            Consensus quality: 178917 bases at least Q30
            Consensus quality: 179096 bases at least Q20
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1333: contig of 1333 bp in length
            *     1334     1433: gap of unknown length
            *     1434    35234: contig of 33801 bp in length
            *    35235    35334: gap of unknown length
            *    35335   179452: contig of 144118 bp in length.
FEATURES             Location/Qualifiers
     source          1. .179452
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="13"
                     /clone="RP24-198D11"
     misc_feature    1. .1333
                     /note="assembly_name:Contig21"
     gap             1334. .1433
                     /estimated_length=unknown
     misc_feature    1434. .35234
                     /note="assembly_name:Contig22"
     gap             35235. .35334
                     /estimated_length=unknown
     misc_feature    35335. .179452
                     /note="assembly_name:Contig23
                     clone_end:T7
                     vector_side:right"
ORIGIN

Alignment Scores:
Pred. No.:              3.19          Length:        179452
Score:                  84.00         Matches:       12
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     14            Gaps:          0

US-09-858-332G-15 (1-12) x AC164547 (1-179452)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     107860 CATAACCATAACCATAACCATAACCATAACCATAAC 107825
```

```
RESULT 67
AC116960
LOCUS       AC116960              182870 bp    DNA     linear   INV 12-MAR-2003
DEFINITION  Dictyostelium discoideum chromosome 2 map
            complement(1004496-821614) strain AX4, complete sequence.
ACCESSION   AC116960
VERSION     AC116960.2  GI:28850330
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 182870)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 182870)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-APR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 182870)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 182870)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On Mar 5, 2003 this sequence version replaced gi:19920015.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES             Location/Qualifiers
     source          1. .182870
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /chromosome="2"
                     /map="complement(1004496-821614)"
     CDS             complement(join(1138. .1508,1613. .2039))
                     /note="GeneID exon scores (in order of location ranges):
                     9.99, 12.71 - GSCJ_ID dd_03471"
                     /codon_start=1
                     /product="similar to Plasmodium falciparum (isolate 3D7).
                     Hypothetical protein"
                     /protein_id="AAO53117.1"
                     /db_xref="GI:28850331"
                     /translation="MSIEVFNYLYFEIKLNLNDGDILLLIVNNLNINSEILEAFFDLL
                     PNQFFESNIESILSISSFKYILKNFINFLISTNNLNLLKKCIFKFSLKDLNNNNNNNN
                     NKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNVLSKVCREKDDLKILKNFIKS
                     LDNFIGDNNEIQQDFSIFNCQTALSFCKNKLNVVKGAALIDKLENLKTNLIFIKYIQI
                     TFNSMDNMSTSSSATFDWLSPLSSPLSNTKISNIPYISSTSTTTTLK"
     CDS             join(2854. .5915,5991. .7260)
                     /note="GeneID exon scores (in order of location ranges):
                     166.05, 67.61 - GSCJ_ID dd_00962"
                     /codon_start=1
                     /product="similar to Mus musculus (Mouse). GABA-A receptor
                     epsilon-like subunit"
                     /protein_id="AAM08494.2"
                     /db_xref="GI:28850332"
```

/translation="MEIDNNEFEISFWKVIHNKYLFDRIINYIQCLEWVNYDDYNQLH
IENRIKFKNINSLKWMINNNQLELLKYKIKINDEFIKITENSLKLLFSNFGFQTSATD
EKLNEILIFLLKYRKRELVPFDLVDIGIKCLNLNVIKFLTSEPYSFQIFQSSLEWVVI
NSNSLFLNELSNLSDTISQTIKIISERNENEKKILDSINQSESNKKFQISNSSLSSTI
TTTLDSKSPFSFGSSFGSSTSPIPPNPTNTLIQPTPSPLASNSSSPISTTFKPTSFGF
STQSQPQPQPQPQPQSQSQSQPQPQSQPTHSATPTIAPFGLSSQPTPTPGSYSSLPGS
TTTTTFRPTSFGFSLISPQPTPISAPIGFSTQALSQPTPTSAPISFSAQSQPQQPQSQ
SQSQSQSQSQPQSQSQPQSQSQPQSQSQPTPTSAPIGFSTQSQPQPQLQSQSQSQFGS
FKAEKVIINPKFISTFSQGSFSNNNFNNNNIDQFIELSEILRSYDLERFKNYLLENKS
NMDNIIIKVNNHKKMMGEKLGNNIDRFCEFMKFYTSNISFNNMDFDRLIQNYIFSTST
NQDEFNKNIEKTISSSNTRLLNLLPSKWNQLLFDTFNFCLDPIGKTYFSIDKKETIDW
IFENLKNQYFIQNLIQSKQVGFLDFSSLEIINYAKNKFNQVSADINSTTTTATIKPIN
NLNEFFTNSEKELSFIYRYSFFKSIGDCNFDQLQFITSLKMNPIECYRENSSNKNGSV
CGSDVRNEKFLKSNLKKQSTENLIKLIEILNNCDYKIFSSELFFKFLSIIGNHCENSE
LLLSKIQFSFTNLDKHFEQYSLNIFMDSLLNNINIIKLFFRNDFNPLKYLNQIKRNGT
TVIPSNFIFSIDHYQPESYNNNHNNILVGILNQSTTESPEMIIEIIKLLLNNFQIDIK
VHKQKVGIQQQKSLDEQDPLQRDMKMVIPHSLLTNFNQYFISLEMIINYLFKKLIQIK
GLPINRLEYIYKIIKNTSVENFSIVLDHSIFHSIYYLAIKSPKLVKYIIIDSCYNIVK
YLDPKIYNFNFIFGELLPQEKEEKEGENIENLILKLILELSHQPVIMSDKSLSNNYYT
TGLNYNLQFYLDIHRIGLFFKQLDFIQSLLDDENFKFGDFPNGIVKNLKIKNIGTCDS
IECCQQQQQQNDNHLKPLLNNRCFHYKDLTIQKASKVMNSQQFFKFIKFCNYHSIFSF
LLIYNDDLLEFETINQLLNQSSIINHLLINHYEPNQDIDYQEEFENQDEIKNQDEIEN
QDEIENQDEIESQDEIENQDEIESQDEINDKDKVEPSEKDDNSDVSEFWLDLIDYCVY
DCEKIEFLFIKHPNLLSFIPIDKILDRVLQPGTNYNYNIVLTDLIVNHINNNNNNQDN
PINCKDILPFFKITNRIIDDAIVAKKIYYLKYIYENGFIEENELLQYLKDKLANNSDY
KNLNWFN"

CDS complement(join(7583. .7606,7699. .7900,8256. .8360,
8496. .8503))
/note="GeneID exon scores (in order of location ranges):
-3.07, 11.66, 5.09, -1.17 - GSCJ_ID dd_00960"
/pseudo
/codon_start=1

CDS complement(join(8558. .8649,8790. .8847))
/note="GeneID exon scores (in order of location ranges):
2.71, 5.62 - GSCJ_ID dd_00996"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAO53118.1"
/db_xref="GI:28850333"
/translation="MELEKVEKDKLELEVVGDYVFDIPSELISFGNESSWLNTNSDKF
KTFLN"

CDS complement(join(9304. .9432,9619. .10184,10260. .12186,
12335. .12463))
/note="GeneID exon scores (in order of location ranges):
2.33, 62.67, 203.85, 8.24 - GSCJ_ID dd_00995"
/codon_start=1
/product="similar to Homo sapiens (Human). DNA mismatch
repair protein Msh2"
/protein_id="AAO53119.1"
/db_xref="GI:28850334"
/translation="MSDNEQEESSQVVLKEDKTFVTFFQSLVSSNEDTDTIRLFDRKK
SLKYWSISDPNPKKKIKIDNDGSLTTTASSSQQQQQELGLAVLTIRQGYEFENIVKEL
LDEKKKIEIWSMKPNSKQQWELIKKGSPGNTQMFEDVLLNGNCEGSVMMALKVTREKG
SIVFGISFGDATFKTIGVSQFMDNDNLSNLSSFIMQMSVKECLLCCDQKNYDYQKVKE
KLSDAGIPFTELPKSDFSSKNAEQDLTRLLGSVKNNLPDIEQEHAIQSASCLIKHLDL
LSNPNYFGKFKLEKYDLDRYMKLDSSSFKGLHIIDLKDSSVSAAAGGGGAGGASSSSN
KDQSLYNLLNQCNTPMGSRLLLQWVKQPLLNAEEIEARLNFVEAFYNDLELRQSLRSN
DLKKIGDLDRLSKKLHGQKATLEDCVNLYGIVTRLPVVLQSLNNHSSIHQELIKVNFI
ESLESIISDFAKFCAMVEKTIDLDLANDKHEYVIRSSFDETLRGIQLKKDQISNKIER
FRVDIADDLNLDEAKVKLHYSEKDMFLLRISRKDEVAIRDKKKYIVHATAKDGVRFAT
REIDTLNEAYKKWSAEYLDKQDGLAKRTLQIAASFVPLIEDLSSLIATLDVFVTLSHV
SSIAPIPFIRPEIIPLGSDENGAGTVIIGGRHPCVEIQDNVNFIANDIDLTRGQSQFQ
IITGPNMGGKSTFIRQVGLIVLMAQIGCFVPAQKATIAVVDCILSRVGAGDSQLRGVS
TFMAEMLETSYILKVATKNSLIIIDELGRGTSTYDGFGLAWGIAEYICNQIGGFCLFA
THFHELTILSDLLPMVKNLHVSASTQNNTFTLLYKVEQGPCDQSFGIHVAILANFPSQ
VIENAKQKAKELESFESNTLKQNHNKFLEEFKEINFNSNDVEKSLSLVNSLLNKYSID
IN"

CDS complement(join(12839. .13000,13400. .13435))
/note="GeneID exon scores (in order of location ranges):
2.76, -2.24 - GSCJ_ID dd_03346"
/codon_start=1
/product="similar to Plasmodium falciparum. Hypothetical
protein"

/protein_id="AAO53120.1"
/db_xref="GI:28850335"
/translation="MLSNGLKIDQEKKFFILIIIIINNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNKNNKNNNKNND"

CDS 14458. .15615
/note="GeneID exon scores (in order of location ranges):
104.42 - GSCJ_ID dd_00993"
/codon_start=1
/product="similar to Mus musculus (Mouse). Similar to
S-adenosylmethionine decarboxylase 1"
/protein_id="AAM08497.2"
/db_xref="GI:28850336"
/translation="MENNNDYKNRSTEIFNEADAGFVYQSNSNDYSDGFEGPEKKLDI
RFGPISKGSVKSIAGSPSKVGLRTIDKEKWQTVLDSARCTIISQTSNDHMDSYVLSES
SLFVYPRRAMIKTCGTTTLLHLIAKMVQVGKECGLEVEMVVFSRKNLNQPSKQVFPHC
SFSDEVNFLNKIFDGQAYVMGDVNKDHWNLYIADFRKNPTLQRTEQTFEVMMHDLDET
VMKQFFKREGVSAWDTTVNSGIADLLPGSMIDDFQFDPCGYSMNGLLNEFYWTIHITP
ESHCSYVSFDTNVALADYNQLLAKVLNVFKPGRFTAALYAEDGAPCGDPYTAFDVNVP
SYAIQNKTVHGFDGGYDVVVSNYQQLDKKVNFSNDDLIQSIESTNIDLIQV"

CDS complement(16030. .17082)
/note="GeneID exon scores (in order of location ranges):
70.61 - GSCJ_ID dd_00992"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAM08498.2"
/db_xref="GI:28850337"
/translation="MKCFILNNIKRKNIPNNIFKASYSHPSFNNSNNIIRNFSSNNKI
NENSNNEKDENLMIPSRSRLSKELESLGAYLPNEEYTEEVEFDDEEYYEDENDDDNNN
NNNNNEIKEEEKEEEKVENQLNEKGIFIKGGKKFQDMIYRIKENHELSNKHSIDLFND
VGEMDSFEEIERKVEERLKKKYQEKMEILDSMDQSDFIYKDNEKVDLYNKISENFIKA
DTHFETPFEITDKQLENNQLIKPNGLSSMISNSDSAFNKVTQSDDISSRISKIKQRIQ
KDKLTNHRINGENLVDLFSLYREDPILNNAQFLAKKFNLSEKSVENLLKYTTIPIIVK
FNTGIKSGHWNVIYET"

CDS join(18407. .18484,18580. .21294)
/note="GeneID exon scores (in order of location ranges):
7.28, 279.86 - GSCJ_ID dd_03068"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAO53121.1"
/db_xref="GI:28850338"
/translation="MDKYDTILWRLRELEENSKCADCTDSFPRYMNTTYGTFVCSVCG
AIHRELGNRVKSISSDKFTQQDIERLEKVGNKMADEIWLSKWSQQQYPLPFPSDEKRV
RDFIKMKYIEKKWIKEGIKQSDFISPPTQSTPGSTPTFTSPKQQQQQQPSAVSQKQQT
TSPLRNSGEKKEILTSKDSLSKEFESLSLGNSNNLLNTNQQQQQQQQQQQQQQQPSTQ
TSQQSIPFIQQAEINQQQQQQQQQFDSNGKPFNPFRDENKYFNSKPQPIESKASFPPF
SEKPFNQHSYLNEKSTMNADQSLGNNSSFYSKIQVQHPTPKSPPTSIYDTQQYKQQQQ
QQQQQQQQQKPQQYQNQQQHQQKLQPQQQSQYQQQQQQQPPPPYINQQPQYNQTNSQM
NQMNGHANQMNNQMNQMNNQMNQMGNSIMDGNQMVNNFQQQQQQPQQYNPFIDTNLSN
NQLALVPVEQNNNSNNNNNQMAVYPQQPPQPQQHNMNFQQQHQQYQPQQQPQQQNLSN
INMNNSNSNIPNGNPFGISDQVVLKHQQMHQQNMNQQNMQQPQMMMNSNGQMVMVTGG
PATTTTTTIIHTPINGNLPSESSSSSGNPFDQQQQQQQQPLDEMFGKKVNSDPVFPAS
SSSFESGNSLDRSDEEQKEAWRRIQEKNIQIQNDKEIARKLQLEEEEIFKTRPRSNTE
PSPRDRTTSQLQLQQSQPQPQQQQYHYSSYSSSYDTANNHGNGMPITQQHIVSGHHRS
SSGSNILKRLSGRSRSKSYDYDDPFEINPRPLNKSGGYDDYDSFMRKNKRIPGGGSNY
QPKISQQDYYQDNNNNNNNNNNNYQQQPSSQQIVVHEQSNYNNNNNNNNANRSYNECD
YCGKDVGLESMSYHKNTECILRQERCSMCNKLVKHMQMGDHQLQCQTNTYTCGVCGRL

Alignment Scores:

Pred. No.:	3.25	Length:	182870
Score:	84.00	Matches:	12
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-15 (1-12) x AC116960 (1-182870)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db       3686 CACAATCACAACCACAACCACAACCACAACCACAAC 3721
```

RESULT 68
AC117176/c

```
LOCUS       AC117176              182871 bp    DNA     linear   INV 12-MAR-2003
DEFINITION  Dictyostelium discoideum chromosome 2 map 5018074-5200947 strain
            AX4, complete sequence.
ACCESSION   AC117176 AC115586 AC115587 AC115596 AC115614 AC116923 AC117071
            AC117178
VERSION     AC117176.2  GI:28829570
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 182871)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 182871)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 182871)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 182871)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On or before Mar 4, 2003 this sequence version replaced
            gi:19569950, gi:20069691, gi:19570012, gi:19909381, gi:19569956,
            gi:20066238, gi:19570143, gi:20069689.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES             Location/Qualifiers
     source          1. .182871
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /chromosome="2"
                     /map="5018074-5200947"
     CDS             201. .4191
                     /note="GeneID exon scores (in order of location ranges):
                     188.00 - GSCJ_ID dd_03459"
                     /pseudo
                     /codon_start=1
     CDS             join(4599. .4898,5000. .5833)
                     /note="GeneID exon scores (in order of location ranges):
                     10.51, 75.05 - GSCJ_ID dd_00370"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO52088.1"
                     /db_xref="GI:28829571"
                     /translation="MFGSKSNFILLIITTLICISLTHCKSFTPSLLSSSSSSFSSSSS
                     SKNKYLYSISSVNPNIESGQIVLVIDPLTNEVIKNKTVYIPYRIIDLLDVDESSNSLL
                     FFCLINDAKNLLLSLNLDTFEFKEIGGTYGGWDYTYARQPYIFVNDNVFLPAKLSFYN
                     ATGLTVLHWDFTDKNSKGGSFKVIEIPIDDFDVESGIPSMGGYDYVNDLLYVTYQSLN
                     DLSENNSTSIVCFNPFSSDTNSTHKIHTGITDLIPGYIHLLFTDQNEGGSLYALSKSF
                     NTQSSIEVCKINFQTNNCDQILSKEMGAYYNSFDYRPYFIGGPQNSDLILLEFTEKNA
                     KPFINFQSVNLLTWNESSPTLLSNYWTSTFGPTYLTFTFFAYI"
     CDS             complement(6175. .6489)
                     /note="GeneID exon scores (in order of location ranges):
                     36.92 - GSCJ_ID dd_00368"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO52089.1"
                     /db_xref="GI:28829572"
                     /translation="MGHHHHHHGHHGHHHHDTITYTTGVYGAPVVPMVQPMVQPMVQP
                     MPIYGQPPMMGAPVMVPPPMMGAPVMVPQPMVYPTTTVVYETGHHHHHGHHHGHHHGH
                     FF"
     CDS             complement(join(7112. .7299,7429. .7809,7910. .8072))
                     /note="GeneID exon scores (in order of location ranges):
                     6.62, 14.60, -0.44 - GSCJ_ID dd_00365"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO52090.1"
                     /db_xref="GI:28829573"
                     /translation="MISSKTTQIKPFYILCIISLIVPITFITTYAITEVRGDDDQIIP
                     YISDTIDFAPESCIGTFGLSIIAFLMIIVVFIKHLIIKNNIIEKEYFDDKKEKKELRF
                     NRFSTSMGFISGLSMHGVSSFQFHNANLVHVIFAALFFLCGFIYLLTQTILDIKNSSD
                     IPLKVLIIKKILICISCLFIGYIITQFFIDNLAAIFEIMSALSIFAYMVTFFYEFSNL
                     TLNIEFSKSIKKINGNGGYHNIESN"
     CDS             9857. .10129
                     /note="GeneID exon scores (in order of location ranges):
                     5.94 - GSCJ_ID dd_00385"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO52091.1"
                     /db_xref="GI:28829574"
                     /translation="MQLFFFFLGNPLGDSIDATNLLLDYKGGCGTRIIQWPNGYKYSF
                     NHSNPVIDNISIESSSFISLGHNFCNSPESIQVYLDGVQIDKSKIK"
     CDS             join(10249. .10614,10675. .10957,11041. .11405,11600. .12025)
                     /note="GeneID exon scores (in order of location ranges):
                     20.49, 7.68, 14.12, 22.22 - GSCJ_ID dd_00382"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO52092.1"
                     /db_xref="GI:28829575"
                     /translation="MPLKINSVPRLKGGSITIIGERLSSQVNNSTIIVKIGQYDCKNV
                     ISSKNEILCNLESVSNVEKVKLVDLPVNISINGIFNENNLLFSFDTPIISDFILPQGE
                     VKLVGHCFGYSQITQINIDNPILPIKKSKLYIIVNGTKSNTIEIDSSFFVKSVPSSPS
                     VNGQTVNFTLFNINTNNFNATPIMTFQDNSSIKSTDYKNSNEYSTHIYSFEIPKSCGP
                     FDQLCKIKFNGKDEENNISISYYPPLVLNSTTISNSSIGGIITIFGNEFYNEIDLITV
                     GKSKCLNSTFINSTSISCFVEPSNFIINQAKQQQQQQQYVNITINGKSGGNNLNIYFDN
                     SFIEKKWAVVYWKYIIKEKIKKFKNRKEQKCIDGMREILNNAEKNFPDIKLDNVITYK
                     KYKMPQEKSVEGELKEQAPPLSEYDQFPSYYLHSPKDSQPSNQIETLPKDSPPPNQNE
                     KLTEDSLPPSNQIETLPEDSSPPPNQTEK"
     CDS             complement(join(12236. .12342,12410. .12826,13073. .13304,
                     13350. .16187))
                     /note="GeneID exon scores (in order of location ranges):
                     8.99, 49.32, 13.83, 125.04 - GSCJ_ID dd_00379"
                     /codon_start=1
                     /product="similar to Plasmodium falciparum. Hypothetical
                     220.3 kDa protein"
                     /protein_id="AAO52093.1"
                     /db_xref="GI:28829576"
                     /translation="MINNETDEFKKENDSINNLDTNKEDIDMESLESKSSRELIKVAK
                     TLKLKANGRKSEIIERIKKFVNDSRNIKQELTKEKLINKFNSSIYQLPIKDIEKHEQL
                     FWRVFKNKMLFKTIFGNLNCKYHSYDDLIYIDWILNNGYLNLLKDKIKRNEPIIGNYY
                     YDGRCPQSEWYLNIFINIQKEYKSNKKETNAPTNNNNNNNNNNKNNNNNNNNNNNNN
                     NNNNNNNKEFFELLFRNYGKYFISFKKVLKELIEVGNFQGFQVLLDNYFNYCKFGPKN
                     GKIKEQSLVNILNLINIHSHLKMFKYFWETYTDKTNCIIKGLLTNLLINQSNGNSHRS
                     NPPKRYKFLNYLFLQSSVVTKNLFDDLIKINIKSTTNITKFTINDLINYCYSISYINN
                     NNNNNNNNNNDNDDLNLIVPLINEIKSKLSNSQLKLQIKSIIPKLINDEDSNNNNYQ
                     QTTANIKKLIDFCEEKSPIRRTIIQQIEIRIVFNLDFDKDIFKAHLPSSPYPREYFLI
                     ISLEYGNYELFKFIVNHYGDSKVLPPIFKNFKSVSNHGEEYPYILFSNCHSKKQRIQF
                     INNAINDHNNNNNNNNNNNNNNNNNNNNNNNNNNNNIIINNKNKEWSNLSVFTIFHLL
                     ISYDDLELLEIAIKGYGGEDSIVNNHCSNIIKVLEKESMMALHQGYLSSFSYFHRHIK
                     SIKMLDYCFHHLNSFFKRVQDMVSWYSLGRIDLLERFEQLLGITNNNNSNSSDNNNNN
                     NNNSNSNIDLLLPFSKNILEENNSNFERSNLIKTAAYIINKPIGIYKYTEDAITHYLE
                     INRNPRLDIDDIITILSKTNNRMEFNDHLNDNYMNSNKIKLLDWMFNFRYEDLISGRF
                     TISKGYYIGLFISNKLDNNNNNSIKSIFSNDKINNKWIFENNYRIPRYPSPYYQVILE
                     LLDWVKKSGDINLLCKLLNEFYFPIFDGTDFEGVIKNGWLRIIQHLFYDQQYRRVLTK
                     KLSNNNSIYTQKELSLLLEHAISWDQFYIVEFFINQMGFKISEAQIVKAIEIRERAKQ
```

```
                VIDLLQDDQRIKEERDKAKTNQNKYVGIGNDNSDFGYGGGSYGGGYDSDSYGSNQRDS
                YGGNQRDSYGGNQRYSYGGNQRETTRRDSFNGRDEGYGNNNNNNNNNNNNNNNSYDSD
                PYSNTRAEYENYSNRAASGERSRPRAASGSGPSPATPNFQSSQQQQPTLID"
     CDS        complement(join(16716. .16960,17218. .17290))
                /note="GeneID exon scores (in order of location ranges):
                7.26, 1.45 - GSCJ_ID dd_00372"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAO52094.1"
                /db_xref="GI:28829577"
                /translation="MINNETDYNNNEEDIDMESLESKSVSIESAKETLFEFLKGVIKN
                GWLRIIQHLFYDQQYRRVLTKKLSSNNSIYTQTELSLLLRHAISWDQFYIAEFFVDQM
                GLK"
     CDS        complement(17816. .18070)
                /note="GeneID exon scores (in order of location ranges):
                26.78 - GSCJ_ID dd_00371"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAO52095.1"
                /db_xref="GI:28829578"
                /translation="MPPHGHHHHGHHGHHEHVTYTTTPIVQPMVQPMPVYGQPPMMVQ
                PQVVMGAPVVYPPTNVVIETSHHHGHHGHHGHHGHHGHFF"
     CDS        complement(18665. .18934)
                /note="GeneID exon scores (in order of location ranges):
                27.69 - GSCJ_ID dd_00369"
                /codon_start=1
                /product="similar to F22H10.2.p [Caenorhabditis elegans]"
                /protein_id="AAO52096.1"
                /db_xref="GI:28829579"
                /translation="MPPHGHHHHHGHHGHHEHITITPVYAPPQVVQPIYAPPPVVQP
                VYVQPPVVYPQATIVLETGHHHGHHHHGHHGHHDHHHHGHHGHH"
     CDS        complement(19591. .19872)
                /note="GeneID exon scores (in order of location ranges):
                34.42 - GSCJ_ID dd_00366"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAO52097.1"
                /db_xref="GI:28829580"
                /translation="MGHHHHHHGHYEEITYTTTPVYGAQVVQPISPVVGIQPIYGQPP
                IVQPIYGQPPVVQPIYGQPPVVYPQTTTIVYETDNHHHGHHHHHHGHFF"
     CDS        complement(join(20614. .22325,22463. .23814,23918. .24106,
                24226. .24464))
                /note="GeneID exon scores (in order of location ranges):
                176.52, 128.57, 13.58, 19.46 - GSCJ_ID dd_00363"
                /codon_start=1
                /product="similar to Dictyostelium discoideum (Slime
                mold). MkpA protein"
                /protein_id="AAO52098.1"
                /db_xref="GI:28829581"
                /translation="MYQNALNHTNSGYGGMKAEEKTNYLNQSVITGYIKGISSYIPSS
                IKSSIKKQIMNSPVTFEEKDTIVGTYFDECFIQGEKHKILINCYNNGFQVWDLNHTDG
                VKELLSCRDGLTKFCKVLTNPLEPDNESSSFYGKRPLLAVVSGEDNPKVTKNMVRIFS
                LQTTELISMYKFKSPIYNVLSNQQIILVVLKERIVGFNANDMSKIKGLSLPCYPSVTS
                LGVIALGSRWIAYTDYESSYSFIGGHHHSSHYFGGGSSNGSGSSGGGGGGIGYSQNRN
```

Alignment Scores:

Pred. No.:	3.25	Length:	182871
Score:	84.00	Matches:	12
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-15 (1-12) x AC117176 (1-182871)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     159843 CACAACCACAACCACAACCACAACCACAACCACAAC 159808
```

```
RESULT 69
AC011405/c
LOCUS       AC011405              184541 bp    DNA     linear   PRI 27-FEB-2002
DEFINITION  Homo sapiens chromosome 5 clone CTB-46B19, complete sequence.
ACCESSION   AC011405
VERSION     AC011405.6  GI:18921282
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 184541)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 184541)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 184541)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-FEB-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 184541)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-FEB-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Feb 26, 2002 this sequence version replaced gi:8576061.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 100% of Sequence;
            Estimated Total Number of Errors is 0.
            NOTE:  BACTERIAL TRANSPOSON excised at 162460
            NOTE:  This insert is not the entire sequence of the clone (entire
            sequence is 195.7kb). It is clipped at the overlap with AC034243.
            The number of bases overlapped is 32318.
FEATURES             Location/Qualifiers
     source          1. .184541
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTB-46B19"
ORIGIN
```

Alignment Scores:

Pred. No.:	3.28	Length:	184541
Score:	84.00	Matches:	12
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	8	Gaps:	0

US-09-858-332G-15 (1-12) x AC011405 (1-184541)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      27855 CATAATCATAATCATAATCATAATCATAATCATAAT 27820
```

```
RESULT 70
AC116475
LOCUS       AC116475              184741 bp    DNA     linear   ROD 11-FEB-2005
DEFINITION  Mus musculus chromosome 8, clone RP23-386N2, complete sequence.
ACCESSION   AC116475
VERSION     AC116475.7  GI:59276174
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
```

```
REFERENCE   1  (bases 1 to 184741)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 8, clone RP23-386N2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 184741)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 184741)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
            Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J.,
            Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
            DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
            Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,
            Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,
            Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
            MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
            Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
            Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JAN-2005) Broad Institute of MIT and Harvard, 320
            Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 184741)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
            Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J.,
            Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
            DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
            Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,
            Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,
            Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
            MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
            Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
            Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-FEB-2005) Broad Institute of MIT and Harvard, 320
            Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 11, 2005 this sequence version replaced gi:57470614.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Broad Institute of MIT and Harvard
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@broad.mit.edu
            -------------- Project Information
                Center project name: L23877
                Center clone name: 386_N_2
            --------------

            Some of the sequence contained within base pairs 144132 to the end
            of the clone was stolen from accession AC134838.
FEATURES             Location/Qualifiers
     source          1. .184741
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="8"
                     /map="8"
                     /clone="RP23-386N2"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     repeat_region   complement(741. .985)
                     /rpt_family="L1_MM"
     repeat_region   1995. .2616
                     /rpt_family="L1_MM"
     repeat_region   2638. .2674
                     /rpt_family="AT_rich"
     repeat_region   2753. .2865
                     /rpt_family="Lx9"
     repeat_region   2857. .3549
                     /rpt_family="Lx2B"
     repeat_region   3567. .3888
                     /rpt_family="Lx9"
     repeat_region   4137. .4172
                     /rpt_family="B3A"
     repeat_region   4300. .4461
                     /rpt_family="B3"
     repeat_region   complement(7048. .7167)
                     /rpt_family="B1_MM"
     repeat_region   complement(7050. .7241)
                     /rpt_family="B4"
     repeat_region   complement(7425. .7779)
                     /rpt_family="Lx5"
     repeat_region   7779. .8058
                     /rpt_family="Lx5"
     repeat_region   complement(8945. .9251)
                     /rpt_family="L1MA5"
     repeat_region   9353. .9373
                     /rpt_family="AT_rich"
     repeat_region   complement(10209. .16999)
                     /rpt_family="L1_MM"
     repeat_region   18265. .18317
                     /rpt_family="(TTG)n"
     repeat_region   complement(18318. .19138)
                     /rpt_family="Lx6"
     repeat_region   19142. .19259
                     /rpt_family="B1_MM"
     repeat_region   20164. .20188
                     /rpt_family="(TG)n"
     repeat_region   complement(20191. .20378)
                     /rpt_family="Lx8"
     repeat_region   20534. .20565
```

```
                     /rpt_family="(TC)n"
     repeat_region   complement(20637. .21066)
                     /rpt_family="Lx6"
     repeat_region   21128. .21285
                     /rpt_family="MLT1H"
     repeat_region   21705. .21806
                     /rpt_family="B3A"
     repeat_region   21925. .21976
                     /rpt_family="(CAAA)n"
     repeat_region   complement(22144. .22753)
                     /rpt_family="L1"
     repeat_region   22914. .22939
                     /rpt_family="(CAAG)n"
     repeat_region   23017. .23036
                     /rpt_family="(TATTG)n"
     repeat_region   23038. .23444
                     /rpt_family="Lx5"
     repeat_region   23558. .23601
                     /rpt_family="(TG)n"
     repeat_region   25980. .26026
                     /rpt_family="(TTTA)n"
     repeat_region   complement(26909. .27005)
                     /rpt_family="MTD"
     repeat_region   27007. .27532
                     /rpt_family="Lx7"
     repeat_region   27593. .27639
                     /rpt_family="(TTTC)n"
     repeat_region   complement(27777. .28034)
                     /rpt_family="MTD"
     repeat_region   complement(28883. .29338)
                     /rpt_family="L1_MM"
     repeat_region   30097. .30372
                     /rpt_family="ORR1B1"
     repeat_region   30373. .30701
                     /rpt_family="Lx6"
     repeat_region   30750. .30784
                     /rpt_family="AT_rich"
     repeat_region   complement(32531. .32650)
                     /rpt_family="B1_MM"
     repeat_region   33432. .33455
                     /rpt_family="AT_rich"
     repeat_region   33527. .33553
                     /rpt_family="(T)n"
     repeat_region   complement(33631. .34125)

Alignment Scores:
Pred. No.:              3.28          Length:         184741
Score:                  84.00         Matches:        12
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     9             Gaps:           0

US-09-858-332G-15 (1-12) x AC116475 (1-184741)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     117181 CATAACCATAACCATAACCATAACCATAACCATAAC 117216

RESULT 71
AC123741/c
LOCUS       AC123741              188453 bp    DNA     linear   ROD 09-JUL-2004
DEFINITION  Mus musculus chromosome 1, clone RP24-113A5, complete sequence.
ACCESSION   AC123741
VERSION     AC123741.9  GI:50080382
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 188453)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 1, clone RP24-113A5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 188453)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
            Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
            Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
            Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 188453)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
            Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J.,
            Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
            DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
            Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,
            Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,
            Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
            MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
            Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
            Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUL-2004) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 188453)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
            Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J.,
            Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
            DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
            Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,
            Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,
            Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
            MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
            Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
```

```
            Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2004) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 9, 2004 this sequence version replaced gi:49745121.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@broad.mit.edu
            -------------- Project Information
                Center project name: L26529
                Center clone name: 113_A_5
            --------------.
FEATURES             Location/Qualifiers
     source          1. .188453
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="1"
                     /map="1"
                     /clone="RP24-113A5"
                     /clone_lib="RPCI-24 Male Mouse BAC"
     misc_feature    1. .4
                     /note="clone_boundary
                     clone_end:SP6
                     site:MboI"
     repeat_region   complement(210. .697)
                     /rpt_family="L1_MM"
     repeat_region   complement(3240. .4210)
                     /rpt_family="Lx8"
     repeat_region   4430. .4474
                     /rpt_family="(CA)n"
     repeat_region   4756. .4796
                     /rpt_family="(TG)n"
     repeat_region   5336. .5378
                     /rpt_family="AT_rich"
     repeat_region   5827. .6424
                     /rpt_family="L1_MM"
     unsure          complement(5962. .6472)
                     /note="single clone coverage"
     repeat_region   6425. .6451
                     /rpt_family="AT_rich"
     repeat_region   complement(7566. .8612)
                     /rpt_family="Lx7"
     repeat_region   complement(8615. .9281)
                     /rpt_family="Lx5"
     repeat_region   9282. .9322
                     /rpt_family="(TG)n"
     repeat_region   complement(9323. .10080)
                     /rpt_family="Lx5"
     repeat_region   complement(10081. .10223)
                     /rpt_family="RMER1B"
     repeat_region   complement(10590. .10672)
                     /rpt_family="RMER1B"
     repeat_region   complement(10977. .11251)
                     /rpt_family="RMER1B"
     repeat_region   complement(11306. .11765)
                     /rpt_family="Lx5"
     repeat_region   complement(11766. .12110)
                     /rpt_family="ORR1C"
     repeat_region   complement(12118. .12269)
                     /rpt_family="Lx5"
     repeat_region   complement(12250. .12437)
                     /rpt_family="Lx5"
     repeat_region   complement(12418. .12554)
                     /rpt_family="L1"
     repeat_region   12555. .12590
                     /rpt_family="(TG)n"
     repeat_region   12590. .12650
                     /rpt_family="(GA)n"
     repeat_region   complement(12591. .12704)
                     /rpt_family="L1"
     repeat_region   12984. .13017
                     /rpt_family="AT_rich"
     repeat_region   13229. .13287
                     /rpt_family="(TATG)n"
     repeat_region   13582. .13620
                     /rpt_family="(CA)n"
     repeat_region   13632. .14105
                     /rpt_family="RLTR11A"
     repeat_region   14110. .14158
                     /rpt_family="(TG)n"
     repeat_region   complement(14277. .14340)
                     /rpt_family="ID4"
     repeat_region   complement(14515. .14888)
                     /rpt_family="MTD"
     repeat_region   15541. .15573
                     /rpt_family="(TG)n"
     repeat_region   complement(15595. .15681)
                     /rpt_family="RSINE1"
     repeat_region   complement(15849. .16220)
                     /rpt_family="MLT1A"
     repeat_region   complement(16331. .16402)
                     /rpt_family="U5"
     repeat_region   16403. .16513
                     /rpt_family="Lx5"
     repeat_region   18093. .18113
                     /rpt_family="AT_rich"
     repeat_region   complement(18359. .18462)
                     /rpt_family="Lx9"
     repeat_region   18463. .18504
                     /rpt_family="AT_rich"
     repeat_region   complement(18935. .19149)
                     /rpt_family="L1F"
     repeat_region   19153. .20136
                     /rpt_family="L1_MM"
     repeat_region   21300. .21644
                     /rpt_family="ORR1D"
     repeat_region   21823. .21860
                     /rpt_family="AT_rich"

Alignment Scores:
Pred. No.:                3.35          Length:          188453
Score:                    84.00         Matches:         12
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       9             Gaps:            0

US-09-858-332G-15 (1-12) x AC123741 (1-188453)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      12558 CACAACCACAACCACAACCACAACCACAACCACAAC 12523

RESULT 72
AC132110/c
LOCUS       AC132110              189130 bp    DNA     linear   ROD 06-FEB-2004
DEFINITION  Mus musculus BAC clone RP24-188M9 from chromosome 18, complete
            sequence.
ACCESSION   AC132110
VERSION     AC132110.2  GI:41386975
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 189130)
  AUTHORS   VanBrunt,A., Haakenson,W., Van Brunt,A., Kozlowicz,A. and
            Bielicki,L.
```

```
  TITLE     The sequence of Mus musculus BAC clone RP24-188M9
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 189130)
  AUTHORS   Wilson,R.
  TITLE     Sequencing of Mus musculus
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 189130)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 189130)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-2004) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   5  (bases 1 to 189130)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-2004) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Jan 29, 2004 this sequence version replaced gi:22539272.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: M_BB0188M09
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone, fosmid clone or direct clone walk sequence.
            Sequence from the Mouse Genome Sequencing Consortium whole genome
            shotgun may have been used to obtain the consensus sequence; and
            the assembly was confirmed by  restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-24 BAC Library has been constructed by Pieter de Jong and
            coworkers (http://www.chori.org) from male C57BL/6J mouse spleen
            and/or brain genomic DNA. The clone and detailed information can be
            obtained from Pieter de Jong and coworkers at http://www.chori.org

            NEIGHBORING SEQUENCE INFORMATION:
            This sequence is the entire insert of the clone. This clone is
            overlapped by AC122826.
FEATURES             Location/Qualifiers
     source          1. .189130
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="18"
                     /map="18"
                     /clone="RP24-188M9"
                     /clone_lib="RPCI-24"
     repeat_region   45. .281
                     /rpt_family="MaLR"
     repeat_region   337. .554
                     /rpt_family="MaLR"
     repeat_region   693. .926
                     /rpt_family="L1"
     repeat_region   930. .1298
                     /rpt_family="MaLR"
     repeat_region   3943. .4217
                     /rpt_family="L1"
     repeat_region   4973. .5118
                     /rpt_family="Alu"
     repeat_region   5934. .6208
                     /rpt_family="B4"
     repeat_region   6333. .6575
                     /rpt_family="MaLR"
     repeat_region   8955. .9233
                     /rpt_family="B4"
     repeat_region   10160. .10306
                     /rpt_family="Alu"
     repeat_region   10603. .10793
                     /rpt_family="B4"
     repeat_region   11606. .11671
                     /rpt_family="ID"
     repeat_region   12296. .12526
                     /rpt_family="B4"
     repeat_region   12556. .12789
                     /rpt_family="MaLR"
     repeat_region   13920. .14098
                     /rpt_family="B4"
     repeat_region   14188. .14347
                     /rpt_family="B4"
     repeat_region   15090. .15137
                     /rpt_family="B4"
     repeat_region   16274. .16418
                     /rpt_family="MER1_type"
     repeat_region   16526. .16784
                     /rpt_family="B4"
     repeat_region   16761. .16844
                     /rpt_family="ERVK"
     repeat_region   16845. .16895
                     /rpt_family="ERV1"
     repeat_region   16867. .16918
                     /rpt_family="ERV1"
     repeat_region   17615. .17876
                     /rpt_family="B4"
     repeat_region   18353. .18428
                     /rpt_family="ID"
     repeat_region   20048. .20159
                     /rpt_family="Alu"
     repeat_region   20524. .20607
                     /rpt_family="MER2_type"
     repeat_region   20670. .20804
                     /rpt_family="Alu"
     repeat_region   20809. .20883
                     /rpt_family="ID"
     repeat_region   21026. .21206
                     /rpt_family="B2"
     repeat_region   21650. .21768
                     /rpt_family="L1"
     repeat_region   23458. .23843
                     /rpt_family="MaLR"
     repeat_region   24068. .24200
                     /rpt_family="MaLR"
     repeat_region   24201. .24624
                     /rpt_family="ERVK"
     repeat_region   24625. .24846
                     /rpt_family="MaLR"
     repeat_region   24905. .25284
                     /rpt_family="MaLR"
     repeat_region   26075. .26216
                     /rpt_family="B4"
     repeat_region   26237. .26420
                     /rpt_family="B4"
     repeat_region   26438. .26514
                     /rpt_family="ID"
```

```
     repeat_region   26612. .26744
                     /rpt_family="MaLR"
     repeat_region   27909. .27978
                     /rpt_family="ID"
     repeat_region   28166. .28297
                     /rpt_family="B4"
     repeat_region   28307. .28546
                     /rpt_family="L1"
     repeat_region   28543. .28670
                     /rpt_family="L1"
     repeat_region   28677. .28727
                     /rpt_family="B4"
     repeat_region   29224. .29365
                     /rpt_family="Alu"
     repeat_region   30617. .30734
                     /rpt_family="Alu"
     repeat_region   32171. .32227
                     /rpt_family="ERV1"
     repeat_region   32445. .32601
                     /rpt_family="B4"
     repeat_region   32667. .32798
                     /rpt_family="B4"
     repeat_region   32880. .32997
                     /rpt_family="B2"
     repeat_region   33034. .33177
                     /rpt_family="B4"
     repeat_region   33118. .33227
                     /rpt_family="Alu"
     repeat_region   33597. .33674
                     /rpt_family="B4"
     repeat_region   34429. .34495
                     /rpt_family="ERV1"
     repeat_region   36865. .36944
                     /rpt_family="ID"
     repeat_region   37114. .37262
                     /rpt_family="B4"
     repeat_region   37725. .37895
                     /rpt_family="B4"
     repeat_region   38211. .38327
                     /rpt_family="Alu"

Alignment Scores:
Pred. No.:                3.36          Length:          189130
Score:                    84.00         Matches:         12
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       9             Gaps:            0

US-09-858-332G-15 (1-12) x AC132110 (1-189130)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      16658 CACAACCACAATCACAACCACAACCACAACCACAAC 16623

RESULT 73
AC134540
LOCUS       AC134540               189505 bp    DNA     linear   ROD 15-MAY-2004
DEFINITION  Mus musculus BAC clone RP24-400M10 from chromosome 18, complete
            sequence.
ACCESSION   AC134540
VERSION     AC134540.3  GI:45237306
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 189505)
  AUTHORS   Ali,J., Bielicki,L., Meyer,R. and Haakenson,W.
  TITLE     The sequence of Mus musculus BAC clone RP24-400M10
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 189505)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 189505)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-OCT-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 189505)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAR-2004) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   5  (bases 1 to 189505)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAY-2004) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Mar 6, 2004 this sequence version replaced gi:23499622.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: M_BB0400M10
            --------------.

            NOTICE:

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone, fosmid clone or direct clone walk sequence.
            Sequence from the Mouse Genome Sequencing Consortium whole genome
            shotgun may have been used to obtain the consensus sequence; and
            the assembly was confirmed by  restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-24 BAC Library has been constructed by Pieter de Jong and
            coworkers (http://www.chori.org) from male C57BL/6J mouse spleen
            and/or brain genomic DNA. The clone and detailed information can be
            obtained from Pieter de Jong and coworkers at http://www.chori.org

            NEIGHBORING SEQUENCE INFORMATION:
            This sequence is the entire insert of the clone. This clone is
            overlapped by AC132110.
FEATURES                Location/Qualifiers
     source             1. .189505
                        /organism="Mus musculus"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:10090"
                        /chromosome="18"
                        /map="18"
                        /clone="RP24-400M10"
                        /clone_lib="RPCI-24"
     misc_feature       77124. .77476
                        /note="Sequence derived from PCR product of project DNA."
     unsure             78150. .78171
                        /note="Sequence derived from one plasmid subclone."
     unsure             90757. .90891
                        /note="Unresolved simple sequence repeat."
     unsure             118118. .118132
                        /note="Sequence derived from one plasmid subclone."
```

```
ORIGIN

Alignment Scores:
Pred. No.:                3.37          Length:        189505
Score:                    84.00         Matches:       12
Percent Similarity:       100.00%       Conservative:  0
Best Local Similarity:    100.00%       Mismatches:    0
Query Match:              100.00%       Indels:        0
DB:                       9             Gaps:          0

US-09-858-332G-15 (1-12) x AC134540 (1-189505)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      28636 CACAACCACAATCACAACCACAACCACAACCACAAC 28671

RESULT 74
AC131664/c
LOCUS       AC131664              189996 bp    DNA     linear   ROD 21-JUN-2005
DEFINITION  Mus musculus BAC clone RP23-274J16 from chromosome 5, complete
            sequence.
ACCESSION   AC131664
VERSION     AC131664.3  GI:65818125
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 189996)
  AUTHORS   Wang,C., Kozlowicz,A., Bielicki,L. and Haglund,K.
  TITLE     The sequence of Mus musculus BAC clone RP23-274J16
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 189996)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 189996)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 189996)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-2005) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   5  (bases 1 to 189996)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUN-2005) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On May 17, 2005 this sequence version replaced gi:30581677.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: M_BA0274J16


            NOTICE:

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e. phred quality
            >=30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone, fosmid clone or direct clone walk sequence.
            Sequence from the Mouse Genome Sequencing Consortium whole genome
            shotgun may have been used to obtain the consensus sequence. The
            assembly was confirmed by restriction digest.
            This finishing standard has slightly changed from the previous
            Human standard. Specifically, standards for regions of low sequence
            complexity (such as dinucleotide repeats and small unit tandem
            repeats) have been relaxed. These regions are very prevalent in the
            mouse genome, and the return on extended finishing efforts is
            minimal.
            If a sequence meets the criteria of the above statement, it needs
            no comments or tags. If the criteria are not met, such as ambiguous
            bases, then the region is duly annotated.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The  BAC Library has been constructed by Kazutoyo Osegawa and
            Minako Tateno in the laboratory of Pieter de Jong
            (http://www.chori.org) from female C57BL/6J mouse kidney and/or
            brain genomic DNA.  The clone and detailed information can be
            obtained from Research Genetics, Inc. (http://www.resgen.com) or
            Pieter de Jong and coworkers at http://www.chori.org

            This sequence is the entire insert of the clone.
FEATURES              Location/Qualifiers
     source           1. .189996
                      /organism="Mus musculus"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:10090"
                      /chromosome="5"
                      /clone="RP23-274J16"
                      /clone_lib="RPCI-23"
     unsure           23646. .23676
                      /note="Unresolved simple sequence repeat."
     unsure           32957. .33005
                      /note="Sequence derived from one plasmid subclone."
     unsure           32995
                      /note="Unresolved bases"
     misc_feature     56313. .56346
                      /note="Sequence derived from PCR product of project DNA"
     misc_feature     56965. .57103
                      /note="Sequence derived from PCR product of project DNA"
     misc_feature     58274. .58306
                      /note="Sequence derived from PCR product of project DNA"
     misc_feature     75933. .76833
                      /note="Sequence derived from PCR product of project DNA"
     unsure           103413. .103856
                      /note="Unresolved simple sequence repeat."
     unsure           122691. .122694
                      /note="Sequence derived from one plasmid subclone."
     misc_feature     123824. .123864
                      /note="Sequence derived from PCR product of project DNA"
ORIGIN

Alignment Scores:
Pred. No.:                3.37          Length:        189996
Score:                    84.00         Matches:       12
Percent Similarity:       100.00%       Conservative:  0
Best Local Similarity:    100.00%       Mismatches:    0
Query Match:              100.00%       Indels:        0
DB:                       9             Gaps:          0

US-09-858-332G-15 (1-12) x AC131664 (1-189996)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      60972 CATAACCATAACCATAACCATAACCATAACCATAAC 60937

RESULT 75
CR450684/c
```

```
LOCUS       CR450684              190345 bp    DNA     linear   HTG 21-MAR-2005
DEFINITION  Danio rerio clone DKEY-60D5, WORKING DRAFT SEQUENCE, 4 unordered
            pieces.
ACCESSION   CR450684
VERSION     CR450684.3  GI:61673446
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 190345)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAR-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Mar 21, 2005 this sequence version replaced gi:47775867.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK60D5
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 188765 bases at least Q40
            Consensus quality: 189075 bases at least Q30
            Consensus quality: 189263 bases at least Q20
            Insert size: 190045; sum-of-contigs
            Insert size: 178551; 2.8% error; agarose-fp
            Quality coverage: 10.01x in Q20 bases; sum-of-contigs Quality
            coverage: 11.00x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3765: contig of 3765 bp in length
            *     3766     3865: gap of 100 bp
            *     3866    23325: contig of 19460 bp in length
            *    23326    23425: gap of 100 bp
            *    23426   103174: contig of 79749 bp in length
            *   103175   103274: gap of 100 bp
            *   103275   190345: contig of 87071 bp in length.
FEATURES             Location/Qualifiers
     source          1. .190345
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-60D5"
                     /clone_lib="DanioKey"
     misc_feature    1. .3765
                     /note="assembly_fragment:00007"
     misc_feature    3866. .23325
                     /note="assembly_fragment:00059.0"
     misc_feature    23426. .103174
                     /note="assembly_fragment:00397"
     misc_feature    103275. .190345
                     /note="assembly_fragment:01514"
ORIGIN

Alignment Scores:
Pred. No.:              3.38          Length:        190345
Score:                  84.00         Matches:       12
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     14            Gaps:          0

US-09-858-332G-15 (1-12) x CR450684 (1-190345)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     159274 CATAATCATAATCATAATCATAATCATAATCATAAT 159239

RESULT 76
AC034243/c
LOCUS       AC034243              191308 bp    DNA     linear   PRI 20-JUL-2001
DEFINITION  Homo sapiens chromosome 5 clone RP1-204O11, complete sequence.
ACCESSION   AC034243
VERSION     AC034243.7  GI:14971193
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 191308)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 191308)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 191308)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 191308)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jul 20, 2001 this sequence version replaced gi:14572133.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.8% of Sequence;
            Estimated Total Number of Errors is 0.2.
            STS Content:
            SHGC-3896 G16896
            SHGC-13050 G16941
            SHGC-14599 G17002
            WI-17966 G22904
            WI-6875 G05741.
FEATURES             Location/Qualifiers
     source          1. .191308
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="RP1-204O11"
ORIGIN

Alignment Scores:
Pred. No.:              3.4           Length:        191308
Score:                  84.00         Matches:       12
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     8             Gaps:          0

US-09-858-332G-15 (1-12) x AC034243 (1-191308)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
```

```
                  ||||||||||||||||||||||||||||||||||||
Db     186843 CATAATCATAATCATAATCATAATCATAATCATAAT 186808

RESULT 77
AC159207/c
LOCUS       AC159207               191824 bp    DNA     linear   ROD 21-JUN-2005
DEFINITION  Mus musculus BAC clone RP23-150H14 from chromosome 13, complete
            sequence.
ACCESSION   AC159207 AC116754
VERSION     AC159207.2  GI:66841714
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 191824)
  AUTHORS   Adams,S., Kozlowicz,A., Cotton,M. and Haglund,K.
  TITLE     The sequence of Mus musculus BAC clone RP23-150H14
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 191824)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-2005) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 191824)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUN-2005) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 191824)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUN-2005) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jun 1, 2005 this sequence version replaced gi:62000937.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: M_BA0150H14
            Drafting center: WIBR

            NOTICE:

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e. phred quality
            >=30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone, fosmid clone or direct clone walk sequence.
            Sequence from the Mouse Genome Sequencing Consortium whole genome
            shotgun may have been used to obtain the consensus sequence. The
            assembly was confirmed by restriction digest.
            This finishing standard has slightly changed from the previous
            Human standard. Specifically, standards for regions of low sequence
            complexity (such as dinucleotide repeats and small unit tandem
            repeats) have been relaxed. These regions are very prevalent in the
            mouse genome, and the return on extended finishing efforts is
            minimal.
            If a sequence meets the criteria of the above statement, it needs
            no comments or tags. If the criteria are not met, such as ambiguous
            bases, then the region is duly annotated.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The  BAC Library has been constructed by Kazutoyo Osegawa and
            Minako Tateno in the laboratory of Pieter de Jong
            (http://www.chori.org) from female C57BL/6J mouse kidney and/or
            brain genomic DNA.  The clone and detailed information can be
            obtained from Research Genetics, Inc. (http://www.resgen.com) or
            Pieter de Jong and coworkers at http://www.chori.org

            This sequence is the entire insert of the clone.
FEATURES             Location/Qualifiers
     source          1. .191824
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="13"
                     /clone="RP23-150H14"
                     /clone_lib="RPCI-23"
     unsure          26263. .26468
                     /note="Sequence derived from one plasmid subclone."
     misc_feature    43054. .43255
                     /note="Sequence derived from PCR product of project DNA"
     misc_feature    43365. .43429
                     /note="Sequence derived from PCR product of project DNA"
     unsure          49122. .49131
                     /note="Sequence derived from one plasmid subclone."
     unsure          49240. .49263
                     /note="Sequence derived from one plasmid subclone."
     unsure          62287. .62470
                     /note="Unresolved simple sequence repeat."
     misc_feature    109983. .110012
                     /note="Sequence derived from PCR product of project DNA"
     unsure          136652. .136735
                     /note="Unresolved simple sequence repeat."
     unsure          137631. .137944
                     /note="Unresolved simple sequence repeat."
ORIGIN

Alignment Scores:
Pred. No.:                3.41          Length:          191824
Score:                    84.00         Matches:         12
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       9             Gaps:            0

US-09-858-332G-15 (1-12) x AC159207 (1-191824)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      49714 CATAACCATAACCATAACCATAACCATAACCATAAC 49679

RESULT 78
ATCHRIV55
LOCUS       ATCHRIV55              194916 bp    DNA     linear   PLN 16-APR-2005
DEFINITION  Arabidopsis thaliana DNA chromosome 4, contig fragment No. 55.
ACCESSION   AL161555
VERSION     AL161555.2  GI:7268934
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 108274)
  AUTHORS   Peters,S.A., van Staveren,M., Dirkse,W., Stiekema,W., Mewes,H.W.,
            Lemcke,K. and Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   2  (bases 86059 to 193760)
  AUTHORS   Van Der Schueren,J., Vandenbussche,F., Chuang,Y.J., Braeken,M.,
            Robben,J., Volckaert,G., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   3  (bases 191405 to 194916)
  AUTHORS   Wedler,H., Wambutt,R., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
```

```
  JOURNAL   Unpublished
REFERENCE   4  (bases 1 to 194916)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
            this fragment has an overlap with ATCHRIV54 at the 5' end and an
            overlap with ATCHRIV56 at the 3' end.
FEATURES             Location/Qualifiers
     source          1. .194916
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /chromosome="4"
                     /ecotype="Columbia"
     gene            7083. .10954
                     /gene="AT4g21380"
     gene            complement(join(7083. .7409,7494. .7644,7741. .7978,
                     8070. .8280,8359. .8552,8642. .8776,9658. .10954))
                     /gene="AT4g21380"
     CDS             complement(join(7083. .7409,7494. .7644,7741. .7978,
                     8070. .8280,8359. .8552,8642. .8776,9658. .10954))
                     /gene="AT4g21380"
                     /note="Contains Protein kinases signatures and profile
                     AA532-554;Protein kinases signatures and profile AA647-659
                     contains EST gb:AI996080.1, AI996128.1, T44910,
                     AI994945.1, AI996494.1, T21059, T43643, T41816,
                     AI992665.1, AI996706.1, T04455"
                     /codon_start=1
                     /product="receptor-like serine/threonine protein kinase
                     ARK3"
                     /protein_id="CAB81245.1"
                     /db_xref="GI:7268935"
                     /db_xref="GOA:O81905"
                     /db_xref="InterPro:IPR000719"
                     /db_xref="InterPro:IPR000858"
                     /db_xref="InterPro:IPR001480"
                     /db_xref="InterPro:IPR003609"
                     /db_xref="InterPro:IPR008271"
                     /db_xref="UniProt/TrEMBL:O81905"
                     /translation="MRGLPNFYHSYTFFFFFLLILFPAYSISANTLSASESLTISSNN
                     TIVSPGNVFELGFFKPGLDSRWYLGIWYKAISKRTYVWVANRDTPLSSSIGTLKISDS
                     NLVVLDQSDTPVWSTNLTGGDVRSPLVAELLDNGNFVLRDSKNSAPDGVLWQSFDFPT
                     DTLLPEMKLGWDAKTGFNRFIRSWKSPDDPSSGDFSFKLETEGFPEIFLWNRESRMYR
                     SGPWNGIRFSGVPEMQPFEYMVFNFTTSKEEVTYSFRITKSDVYSRLSISSSGLLQRF
                     TWIETAQNWNQFWYAPKDQCDEYKECGVYGYCDSNTSPVCNCIKGFKPRNPQVWGLRD
                     GSDGCVRKTLLSCGGGDGFVRLKKMKLPDTTTASVDRGIGVKECEQKCLRDCNCTAFA
                     NTDIRGSGSGCVTWTGELFDIRNYAKGGQDLYVRLAATDLEDKRNRSAKIIGSSIGVS
                     VLLLLSFIIFFLWKRKQKRSILIETPIVDHQLRSRDLLMNEVVISSRRHISRENNTDD
                     LELPLMEFEEVAMATNNFSNANKLGQGGFGIVYKGKLLDGQEMAVKRLSKTSVQGTDE
                     FKNEVKLIARLQHINLVRLLACCVDAGEKMLIYEYLENLSLDSHLFDKSRNSKLNWQM
                     RFDIINGIARGLLYLHQDSRFRIIHRDLKASNILLDKYMTPKISDFGMARIFGRDETE
                     ANTRKVVGTYGYMSPEYAMDGIFSMKSDVFSFGVLLLEIISSKRNKGFYNSDRDLNLL
                     GCVWRNWKEGKGLEIIDPIITDSSSTFRQHEILRCIQIGLLCVQERAEDRPTMSLVIL
                     MLGSESTTIPQPKAPGYCLERSLLDTDSSSSKQRDDESWTVNQITVSVLDAR"
     exon            complement(7083. .7409)
                     /gene="AT4g21380"
                     /number=1
     intron          complement(7410. .7493)
                     /gene="AT4g21380"
                     /number=1
     exon            complement(7494. .7644)
                     /gene="AT4g21380"
                     /number=2
     intron          complement(7645. .7740)
                     /gene="AT4g21380"
                     /number=2
     exon            complement(7741. .7978)
                     /gene="AT4g21380"
                     /number=3
     intron          complement(7979. .8069)
                     /gene="AT4g21380"
                     /number=3
     exon            complement(8070. .8280)
                     /gene="AT4g21380"
                     /number=4
     intron          complement(8281. .8358)
                     /gene="AT4g21380"
                     /number=4
     exon            complement(8359. .8552)
                     /gene="AT4g21380"
                     /number=5
     intron          complement(8553. .8641)
                     /gene="AT4g21380"
                     /number=5
     exon            complement(8642. .8776)
                     /gene="AT4g21380"
                     /number=6
     intron          complement(8777. .9657)
                     /gene="AT4g21380"
                     /number=6
     exon            complement(9658. .10954)
                     /gene="AT4g21380"
                     /number=7
     gene            12322. .15338
                     /gene="AT4g21390"
     gene            complement(join(12322. .12639,12758. .12905,13005. .13242,
                     13324. .13534,13619. .13854,13940. .15338))
                     /gene="AT4g21390"
     CDS             complement(join(12322. .12639,12758. .12905,13005. .13242,
                     13324. .13534,13619. .13854,13940. .15338))
                     /gene="AT4g21390"
                     /note="strong similarity to serine/threonine kinase BRLK,
                     Brassica oleracea, gb:Y12531
                     Contains Prokaryotic membrane lipoprotein lipid attachment
                     site AA392-402;Protein kinases signatures and profile
                     AA650-662
                     contains EST gb:T04455, T43643, AI992665.1, T41816,
                     AI996080.1, T21059, AI996706.1, AU036593.1, T21145,
                     R65507, AI993345.1"
                     /codon_start=1
                     /product="serine/threonine kinase-like protein"
                     /protein_id="CAB81246.1"
                     /db_xref="GI:7268936"
                     /db_xref="GOA:O81906"
                     /db_xref="InterPro:IPR000719"
                     /db_xref="InterPro:IPR000858"
                     /db_xref="InterPro:IPR001480"
                     /db_xref="InterPro:IPR003609"
                     /db_xref="InterPro:IPR008271"
                     /db_xref="UniProt/TrEMBL:O81906"
                     /translation="MRFFRKTSLYLSLFLYFFLYESSMAANTIRRGESLRDGINHKPL
                     VSPQKTFELGFFSPGSSTHRFLGIWYGNIEDKAVVWVANRATPISDQSGVLMISNDGN
                     LVLLDGKNITVWSSNIESSTTNNNNRVVSIHDTGNFVLSETDTDRPIWESFNHPTDTF
                     LPQMRVRVNPQTGDNHAFVSWRSETDPSPGNYSLGVDPSGAPEIVLWEGNKTRKWRSG
                     QWNSAIFTGIPNMSLLTNYLYGFKLSSPPDETGSVYFTYVPSDPSVLLRFKVLYNGTE
                     EELRWNETLKKWTKFQSEPDSECDQYNRCGKFGICDMKGSNGICSCIHGYEQVSVGNW
                     SRGCRRRTPLKCERNISVGEDEFLTLKSVKLPDFEIPEHNLVDPEDCRERCLRNCSCN
                     AYSLVGGIGCMIWNQDLVDLQQFEAGGSSLHIRLADSEVGENRKTKIAVIVAVLVGVI
                     LIGIFALLLWRFKRKKDVSGAYCGKNTDTSVVVADLTKSKETTSAFSGSVDIMIEGKA
                     VNTSELPVFSLNAIAIATNDFCKENELGRGGFGPVYKGVLEDGREIAVKRLSGKSGQG
                     VDEFKNEIILIAKLQHRNLVRLLGCCFEGEEKMLVYEYMPNKSLDFFLFDETKQALID
                     WKLRFSIIEGIARGLLYLHRDSRLRIIHRDLKVSNVLLDAEMNPKISDFGMARIFGGN
                     QNEANTVRVVGTYGYMSPEYAMEGLFSVKSDVYSFGVLLLEIVSGKRNTSLRSSEHGS
                     LIGYAWYLYTHGRSEELVDPKIRVTCSKREALRCIHVAMLCVQDSAAERPNMASVLLM
                     LESDTATLAAPRQPTFTSTRRNSIDVNFALDSSQQYIVSSNEITSTVVLGR"
     exon            complement(12322. .12639)
                     /gene="AT4g21390"
                     /number=1
```

```
     intron          complement(12640. .12757)
                     /gene="AT4g21390"
                     /number=1
     exon            complement(12758. .12905)
                     /gene="AT4g21390"
                     /number=2
     intron          complement(12906. .13004)
                     /gene="AT4g21390"
                     /number=2
     exon            complement(13005. .13242)
                     /gene="AT4g21390"
                     /number=3
     intron          complement(13243. .13323)
                     /gene="AT4g21390"
                     /number=3
     exon            complement(13324. .13534)
                     /gene="AT4g21390"
                     /number=4
     intron          complement(13535. .13618)
                     /gene="AT4g21390"
                     /number=4
     exon            complement(13619. .13854)
                     /gene="AT4g21390"
                     /number=5
     intron          complement(13855. .13939)
                     /gene="AT4g21390"
                     /number=5
     exon            complement(13940. .15338)
                     /gene="AT4g21390"
                     /number=6
     gene            17082. .19573
                     /gene="AT4g21400"
     gene            complement(join(17082. .17390,17476. .17632,17702. .18240,
                     18325. .18437,18489. .18665,18733. .19573))

Alignment Scores:
Pred. No.:              3.46          Length:         194916
Score:                  84.00         Matches:        12
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     15            Gaps:           0

US-09-858-332G-15 (1-12) x ATCHRIV55 (1-194916)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     194880 CACAATCACAATCACAATCACAATCACAATCACAAT 194915

RESULT 79
ATCHRIV56
LOCUS       ATCHRIV56             196247 bp    DNA     linear   PLN 16-APR-2005
DEFINITION  Arabidopsis thaliana DNA chromosome 4, contig fragment No. 56.
ACCESSION   AL161556
VERSION     AL161556.2  GI:7269026
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1489 to 157589)
  AUTHORS   Wedler,H., Wambutt,R., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   2  (bases 157343 to 167411)
  AUTHORS   Robben,J., Grymonprez,B., Volckaert,G., Mewes,H.W., Lemcke,K. and
            Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   3  (bases 167197 to 196247)
  AUTHORS   Murphy,G., Ridley,P., Hudson,S., Mewes,H.W., Lemcke,K. and
            Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   4  (bases 1 to 196247)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
            this fragment has an overlap with ATCHRIV55 at the 5' end and an
            overlap with ATCHRIV57 at the 3' end.
FEATURES             Location/Qualifiers
     source          1. .196247
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /chromosome="4"
                     /ecotype="Columbia"
     gene            5510. .11823
                     /gene="AT4g21820"
     CDS             join(5510. .6217,6310. .6778,7223. .7449,7548. .7613,
                     7799. .7852,8220. .8336,8544. .8642,8699. .8879,9087. .9224,
                     9426. .9620,9702. .9982,10872. .11094,11161. .11425,
                     11469. .11624,11736. .11823)
                     /gene="AT4g21820"
                     /note="similarity to calmodulin-binding protein SHA1, Mus
                     musculus, AF062378
                     contains EST gb:AI997648.1"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB79137.1"
                     /db_xref="GI:7269027"
                     /db_xref="InterPro:IPR000048"
                     /db_xref="InterPro:IPR000225"
                     /db_xref="UniProt/TrEMBL:O49705"
                     /translation="MDENEPPCASPAPPRNPASSLLSDISNFKTPRRTSVVNSNISKS
                     PYPHFFTASKQTPKSSSSNFRRPSMVHSYASRSKVSTSSRRLKAFELQQSQSSRKAEL
                     TKEKNLRSLAKSLTVWLNFLFENPENCGCDPFENESGVGNLGKGKRDSGEALGNSKSV
                     GVDTMWRSPKRLRNLGWCGEKKRSEIDSSLTGSKYSTLRESLRDVCSLDDLKQRMQFH
                     LSLGSCKEIFDVMTRVSKNIDEGRIKMKPQCPLVTDFGMKEKAIKALMSYNQVWLRLG
                     LYIIFGGDSFLSDSEVNSDQEMAFLKMVISKQFFSHDGLARSYVYNKMVEGLYRPGYY
                     EALGNVILKRILLLVLVIDRAKSQSCLSLKYGIDGIDGGSPLMFSEKSSIKSSHQLIC
                     GSEFDLMRVLIKQKMVVPSDNRKKNLANCRIALQYLKDAGVSLKDDEGMMITGEDVAD
                     GDRELTISLLWNIFVHLQLPLLINGRLLTEEIYKVQGVEQNNQITMSTPLEMLLNWIQ
                     DPGGQEGPQSVMSNTDYHDAVQNFILSQKLTALLGSFPEIGDLLEHNAVVSNQSVIIL
                     LAFLSSKLIVKENLLDFTLAGYLTNSVFSQEQLNFHKLLCSSCQVQEKRYSRIRISCS
                     SSEAVTIEEPDRENGEDATKRFQAIKAWWQDMANQNQISVGKANSHTLQGSLSRKCTT
                     DSQRAYLLLAEIAAVIIQSNIRGLHARRKFRNKMRAICFLQAAVRTWLSVKNIQVVEK
                     FNVEEVTLHLSERSANLKPVARYVKFIVDRSRFIKLRKSVSVIQKAVRRHQSNLHHEL
                     KAALRIQLAWRSYKEKVISSITIQSYVRGWITRRMNRTYKFSSILIQFQVSFYPYISQ
                     SHWRGYLTRKASKAQVLDLRVRMQTSAANIDDKKRLINKLLSALSELLSMKNVHNILH
                     ICETLESATKYSDKCCEELVAAGAIEKLLTLIRSASRSIPDQQVSKHALSTLGHLARY
                     PQMADELINTKGSIQTIFWELLRSSFSSQVFFLSFKNKEEAYFIASDVLKKICSSHKG
                     VEAVRKLPALVKRLHVLVEELTRKANIEKRNVKGQTGKEKSERRLKEAIELVKLITRR
                     "
     exon            5510. .6217
                     /gene="AT4g21820"
                     /number=1
     intron          6218. .6309
                     /gene="AT4g21820"
                     /number=1
     exon            6310. .6778
                     /gene="AT4g21820"
                     /number=2
     intron          6779. .7222
                     /gene="AT4g21820"
                     /number=2
     exon            7223. .7449
                     /gene="AT4g21820"
                     /number=3
     intron          7450. .7547
```

```
                /gene="AT4g21820"
                /number=3
     exon       7548. .7613
                /gene="AT4g21820"
                /number=4
     intron     7614. .7798
                /gene="AT4g21820"
                /number=4
     exon       7799. .7852
                /gene="AT4g21820"
                /number=5
     intron     7853. .8219
                /gene="AT4g21820"
                /number=5
     exon       8220. .8336
                /gene="AT4g21820"
                /number=6
     intron     8337. .8543
                /gene="AT4g21820"
                /number=6
     exon       8544. .8642
                /gene="AT4g21820"
                /number=7
     intron     8643. .8698
                /gene="AT4g21820"
                /number=7
     exon       8699. .8879
                /gene="AT4g21820"
                /number=8
     intron     8880. .9086
                /gene="AT4g21820"
                /number=8
     exon       9087. .9224
                /gene="AT4g21820"
                /number=9
     intron     9225. .9425
                /gene="AT4g21820"
                /number=9
     exon       9426. .9620
                /gene="AT4g21820"
                /number=10
     intron     9621. .9701
                /gene="AT4g21820"
                /number=10
     exon       9702. .9982
                /gene="AT4g21820"
                /number=11
     intron     9983. .10871
                /gene="AT4g21820"
                /number=11
     exon       10872. .11094
                /gene="AT4g21820"
                /number=12
     intron     11095. .11160
                /gene="AT4g21820"
                /number=12
     exon       11161. .11425
                /gene="AT4g21820"
                /number=13
     intron     11426. .11468
                /gene="AT4g21820"
                /number=13
     exon       11469. .11624
                /gene="AT4g21820"
                /number=14
     intron     11625. .11735
                /gene="AT4g21820"
                /number=14
     exon       11736. .11823
                /gene="AT4g21820"
                /number=15
     gene       12517. .13199
                /gene="AT4g21830"
     gene       complement(join(12517. .12675,12834. .12990,13090. .13199))
                /gene="AT4g21830"
     CDS        complement(join(12517. .12675,12834. .12990,13090. .13199))
                /gene="AT4g21830"
                /note="strong similarity to CGI-131 protein, Homo sapiens,
                AF151889
                contains EST gb:N38012, AA713335, AI996517.1, AA712490,
                H76478"
                /codon_start=1
                /product="putative protein"
                /protein_id="CAB79138.1"
                /db_xref="GI:7269028"
                /db_xref="GOA:O49706"
                /db_xref="InterPro:IPR002579"
                /db_xref="UniProt/TrEMBL:O49706"
                /translation="MTAAAVPATGSFQKQDEEWRAVLSPEQFRVLRLKGTDKRGKGEF
                TKKFEEGTYSCAGCGTALYKSTTKFDSGCGWPAFFDAIPGAIKQTPEAGGRRMEITCA
                VCDGHLGHVFKGEGYSTPTDQRHCVNSVSLKFSSAGSSQ"
     exon       complement(12517. .12675)
                /gene="AT4g21830"
                /number=1
     intron     complement(12676. .12833)
                /gene="AT4g21830"
                /number=1
     exon       complement(12834. .12990)
                /gene="AT4g21830"
                /number=2
     intron     complement(12991. .13089)
                /gene="AT4g21830"
                /number=2
     exon       complement(13090. .13199)
                /gene="AT4g21830"

Alignment Scores:
Pred. No.:                3.48          Length:          196247
Score:                    84.00         Matches:         12
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       15            Gaps:            0

US-09-858-332G-15 (1-12) x ATCHRIV56 (1-196247)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db       4964 CACAATCACAATCACAATCACAATCACAATCACAAT 4999

RESULT 80
AC104863/c
LOCUS       AC104863               198321 bp    DNA     linear   ROD 30-SEP-2003
DEFINITION  Mus musculus chromosome 6, clone RP23-109C11, complete sequence.
ACCESSION   AC104863
VERSION     AC104863.12  GI:30141979
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 198321)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 6, clone RP23-109C11
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 198321)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
```

Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (22-DEC-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 3 (bases 1 to 198321)

AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (23-APR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 4 (bases 1 to 198321)

AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (26-APR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 5 (bases 1 to 198321)

AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Lui,X., Mabbitt,R., MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (30-SEP-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Apr 26, 2003 this sequence version replaced gi:30065658.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-------------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-------------- Project Information
Center project name: L16460
Center clone name: 109_C_11
--------------.

FEATURES Location/Qualifiers

source	1. .198321
	/organism="Mus musculus"
	/mol_type="genomic DNA"
	/db_xref="taxon:10090"
	/chromosome="6"
	/map="6"
	/clone="RP23-109C11"
	/clone_lib="RPCI-23 Female Mouse BAC"
misc_feature	1. .6
	/note="clone_boundary
	clone_end:SP6 end
	site:EcoRI"
repeat_region	complement(21. .818)
	/rpt_family="L1"
repeat_region	820. .7113
	/rpt_family="L1_MM"
unsure	2599. .2626
	/note="single clone coverage"
unsure	2663
	/note="probably T, possibly TT"
repeat_region	complement(7115. .10490)
	/rpt_family="L1"
repeat_region	11031. .11149
	/rpt_family="CT-rich"
repeat_region	11530. .11586
	/rpt_family="(TAGA)n"
repeat_region	11749. .11776
	/rpt_family="AT_rich"
repeat_region	12863. .12926
	/rpt_family="MIR"
repeat_region	14087. .14121
	/rpt_family="(TG)n"
repeat_region	complement(14218. .14383)
	/rpt_family="B4A"
repeat_region	15823. .16059
	/rpt_family="B4A"
repeat_region	complement(16475. .16765)
	/rpt_family="RMER19B"
repeat_region	complement(16883. .17078)
	/rpt_family="L1_MM"
repeat_region	complement(17079. .17461)
	/rpt_family="RMER19B"
repeat_region	18280. .18350
	/rpt_family="ORR1C"

```
     repeat_region   18484. .18580
                     /rpt_family="MTE"
     repeat_region   18586. .18640
                     /rpt_family="(TG)n"
     repeat_region   18927. .19362
                     /rpt_family="RMER6A"
     repeat_region   19623. .20015
                     /rpt_family="L1M4"
     repeat_region   20900. .20952
                     /rpt_family="(TG)n"
     repeat_region   21205. .21256
                     /rpt_family="ID_B1"
     repeat_region   23070. .23097
                     /rpt_family="AT_rich"
     repeat_region   23963. .23986
                     /rpt_family="(T)n"
     repeat_region   complement(25045. .25223)
                     /rpt_family="B4A"
     repeat_region   complement(25293. .25414)
                     /rpt_family="RMER1B"
     repeat_region   25417. .25585
                     /rpt_family="(GAA)n"
     repeat_region   26004. .26237
                     /rpt_family="MLT1G1"
     repeat_region   26562. .26722
                     /rpt_family="MLT1G"
     repeat_region   complement(27250. .27583)
                     /rpt_family="IAPLTR1_MM"
     repeat_region   complement(27586. .28359)
                     /rpt_family="IAPLTR1_MM-int"

Alignment Scores:
Pred. No.:                3.52          Length:          198321
Score:                    84.00         Matches:         12
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       9             Gaps:            0

US-09-858-332G-15 (1-12) x AC104863 (1-198321)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      74600 CACAACCACAACCACAACCACAACCACAACCACAAC 74565

RESULT 81
AC122454
LOCUS       AC122454               198341 bp    DNA     linear   ROD 13-NOV-2003
DEFINITION  Mus musculus BAC clone RP24-263I14 from chromosome 17, complete
            sequence.
ACCESSION   AC122454
VERSION     AC122454.4  GI:24476149
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 198341)
  AUTHORS   Wang,C., Cotton,M. and Schatzkamer,K.
  TITLE     The sequence of Mus musculus BAC clone RP24-263I14
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 198341)
  AUTHORS   Wilson,R.
  TITLE     Sequencing of Mus musculus
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 198341)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 198341)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   5  (bases 1 to 198341)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-NOV-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   6  (bases 1 to 198341)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Nov 2, 2002 this sequence version replaced gi:22476139.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: M_BB0263I14
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-24 BAC Library has been constructed by Pieter de Jong and
            coworkers (http://www.chori.org) from male C57BL/6J mouse spleen
            and/or brain genomic DNA. The clone and detailed information can be
            obtained from Pieter de Jong and coworkers at http://www.chori.org

            NEIGHBORING SEQUENCE INFORMATION:
            This sequence is the entire insert of the clone.
FEATURES                Location/Qualifiers
     source             1. .198341
                        /organism="Mus musculus"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:10090"
                        /chromosome="17"
                        /map="17"
                        /clone="RP24-263I14"
                        /clone_lib="RPCI-24"
     repeat_region      111. .225
                        /rpt_family="Alu"
     repeat_region      535. .642
                        /rpt_family="B4"
     repeat_region      552. .656
                        /rpt_family="Alu"
     repeat_region      859. .873
                        /rpt_family="B2"
     repeat_region      874. .1020
                        /rpt_family="Alu"
     repeat_region      1021. .1099
                        /rpt_family="B2"
     repeat_region      1153. .1327
                        /rpt_family="B2"
```

```
repeat_region   1487. .1604
                /rpt_family="MaLR"
repeat_region   1921. .2023
                /rpt_family="Alu"
repeat_region   2120. .2259
                /rpt_family="L1"
repeat_region   2279. .2375
                /rpt_family="Alu"
repeat_region   2415. .2635
                /rpt_family="MaLR"
repeat_region   2636. .2966
                /rpt_family="MaLR"
repeat_region   2967. .3027
                /rpt_family="MaLR"
repeat_region   3028. .3221
                /rpt_family="B2"
repeat_region   3310. .3358
                /rpt_family="B4"
repeat_region   3381. .3496
                /rpt_family="B4"
repeat_region   3605. .3753
                /rpt_family="Alu"
repeat_region   4396. .4422
                /rpt_family="B2"
repeat_region   4423. .4606
                /rpt_family="B2"
repeat_region   4607. .4770
                /rpt_family="B2"
repeat_region   6602. .7046
                /rpt_family="ERVK"
repeat_region   7521. .7683
                /rpt_family="B2"
repeat_region   8878. .8923
                /rpt_family="ID"
repeat_region   9046. .9101
                /rpt_family="Alu"
repeat_region   11754. .11883
                /rpt_family="RMER14"
repeat_region   11952. .12201
                /rpt_family="ERV1"
repeat_region   12683. .12748
                /rpt_family="ID"
repeat_region   12756. .13101
                /rpt_family="MaLR"
repeat_region   14642. .14771
                /rpt_family="Alu"
repeat_region   14648. .14848
                /rpt_family="B4"
repeat_region   15058. .15248
                /rpt_family="MaLR"
repeat_region   15131. .15249
                /rpt_family="Alu"
repeat_region   15250. .15296
                /rpt_family="B4"
repeat_region   15439. .15615
                /rpt_family="B4"
repeat_region   23433. .23621
                /rpt_family="B4"
repeat_region   24052. .24239
                /rpt_family="B2"
repeat_region   24261. .24361
                /rpt_family="Alu"
repeat_region   24344. .24619
                /rpt_family="B4"
repeat_region   24641. .24861
                /rpt_family="B2"
repeat_region   24876. .25012
                /rpt_family="L1"
repeat_region   25180. .25282
                /rpt_family="Alu"
repeat_region   25737. .25908
                /rpt_family="L1"
repeat_region   25779. .25914
                /rpt_family="L1"
repeat_region   25918. .26347
                /rpt_family="L1"
repeat_region   26638. .26768
                /rpt_family="Alu"
repeat_region   29289. .29419
                /rpt_family="Alu"
repeat_region   36950. .37074
                /rpt_family="MIR"
repeat_region   37091. .37281
                /rpt_family="B2"
repeat_region   44383. .44475
                /rpt_family="Alu"
repeat_region   45853. .46029
                /rpt_family="B4"
repeat_region   46958. .47183
                /rpt_family="MaLR"
repeat_region   47489. .47522
                /rpt_family="B2"
repeat_region   52152. .52291
                /rpt_family="Alu"
repeat_region   53004. .53079
                /rpt_family="ID"
repeat_region   53108. .53205
                /rpt_family="Alu"
repeat_region   53218. .53280
                /rpt_family="B2"
repeat_region   53342. .53413

Alignment Scores:
Pred. No.:              3.52          Length:          198341
Score:                  84.00         Matches:         12
Percent Similarity:     100.00%       Conservative:    0
Best Local Similarity:  100.00%       Mismatches:      0
Query Match:            100.00%       Indels:          0
DB:                     9             Gaps:            0

US-09-858-332G-15 (1-12) x AC122454 (1-198341)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     103983 CACAACCACAACCACAACCACAACCACAACCACAAC 104018

RESULT 82
BX897729
LOCUS       BX897729              201932 bp    DNA     linear   VRT 17-APR-2005
DEFINITION  Zebrafish DNA sequence from clone DKEY-240A12 in linkage group 4
            Contains the gene for a novel protein similar to vertebrate
            neuronal cell adhesion molecule (NRCAM), the gene for a novel
            similar to contactin 1 (cntn1) and the 5' end of a novel gene,
            complete sequence.
ACCESSION   BX897729
VERSION     BX897729.6  GI:49614058
KEYWORDS    HTG.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 201932)
  AUTHORS   Garnett,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-DEC-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests:
            http://www.sanger.ac.uk/Projects/D_rerio/faqs.shtml#dataeight
COMMENT     On Jul 1, 2004 this sequence version replaced gi:49457619.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            --------------
```

```
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived
            Zebrafish pUC subclones occasionally display inconsistency over the
            length of mononucleotide A/T runs and conserved TA repeats. Where
            this is found the longest good quality representation will be
            submitted.
            DKEY-240A12 is from a Zebrafish BAC library
            VECTOR: pIndigoBAC-5
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
            Any regions longer than 1kb tagged as misc-feature 'unsure' are
            part of a tandem repeat of more than 10kb in length where it has
            not been possible to anchor the base differences between repeat
            copies.  The region has been built up based on the repeat element
            to match the total size of repeat indicated by restriction digest,
            but repeat copies may not be in the correct order and the usual
            finishing criteria may not apply.
FEATURES                Location/Qualifiers
     source             1. .201932
                        /organism="Danio rerio"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:7955"
                        /chromosome="4"
                        /clone="DKEY-240A12"
                        /clone_lib="DanioKey"
     misc_feature       14192
                        /note="Tandem repeat. Forced join. Assembly confirmed by
                        restriction digest data."
     gene               38709. .70469
                        /locus_tag="DKEY-240A12.1-001"
     mRNA               join(38709. .38840,42628. .42733,44452. .44645,44749. .44871,
                        45115. .45285,45846. .45905,47800. .47911,48193. .48377,
                        48879. .49010,49624. .49767,49852. .49963,50532. .50695,
                        50836. .50983,51620. .51744,52630. .52731,52817. .53014,
                        54947. .55017,55125. .55347,56303. .56418,56512. .56716,
                        56800. .56922,58583. .58759,60461. .60496,63799. .63921,
                        64723. .64878,65500. .65631,66665. .66743,70233. .70469)
                        /locus_tag="DKEY-240A12.1-001"
                        /product="novel protein similar to vertebrate neuronal
                        cell adhesion molecule (NRCAM)"
                        /note="match: ESTs: Em:AL925548.1 Em:CK238930.1
                        Em:CK317268.1"
     CDS                join(38735. .38840,42628. .42733,44452. .44645,44749. .44871,
                        45115. .45285,45846. .45905,47800. .47911,48193. .48377,
                        48879. .49010,49624. .49767,49852. .49963,50532. .50695,
                        50836. .50983,51620. .51744,52630. .52731,52817. .53014,
                        54947. .55017,55125. .55347,56303. .56418,56512. .56716,
                        56800. .56922,58583. .58759,60461. .60496,63799. .63921,
                        64723. .64878,65500. .65631,66665. .66743,70233. .70467)
                        /locus_tag="DKEY-240A12.1-001"
                        /standard_name="OTTDARP00000008714"
                        /note="match: proteins: Sw:P35331 Sw:P97685 Sw:Q810U4
                        Sw:Q92823 Tr:Q7SXQ5 Tr:Q9QVN3"
                        /codon_start=1
                        /product="novel protein similar to vertebrate neuronal
                        cell adhesion molecule (NRCAM)"
                        /protein_id="CAH69053.1"
                        /db_xref="GI:55251198"
                        /db_xref="GOA:Q5TYT1"
                        /db_xref="InterPro:IPR003006"
                        /db_xref="InterPro:IPR003598"
                        /db_xref="InterPro:IPR003599"
                        /db_xref="InterPro:IPR003961"
                        /db_xref="InterPro:IPR003962"
                        /db_xref="InterPro:IPR007110"
                        /db_xref="UniProt/TrEMBL:Q5TYT1"
                        /translation="MDGKRSCSLCGGAMMIMMMMMISHMTSALEVPLDLPQPPTITN
                        QSPKDYIIDPRENINIFCEAKGKPHPSFSWTRNGTHFDVEKDPKVVMMPGTGNLVIDI
                        SGEKAEVYEGVYQCIARNEHGSAVSNNIVIRQSRSPLWSKEKNEPITVQRGTSLILQC
                        RPPAGLPPPIIFWMDNNFQRLPQNSRVSQALNGDLYFSNVIMEDTRNDYICYARFPHT
                        QTIQQKQPITVKVLDIEAMNDTVLAAFLNGSDFWGDSPSGERVPSFLHPPGMESTTMV
                        LKGDTLELECIADGLPTPNISWTKVNGELPSGRFSFYSFQKTLKIKEVTEADGGDYRC
                        IAKNRMGSSQHTITVVVRAAPFWISAPQNLILAPKETGNLTCDVDGNPKPTITWSVNG
                        NPIESSHNDPSRKVSDGVITLSDVQTGSSAVYQCNASNDYGYLLANAFVSVLAEPPRV
                        LTSLNHEYSVISNSRALLDCASFGSPLPKITWFKDSQSILNSDLYHIHKNGTLEINVA
                        QPQNSGKYTCIASNNLGNKENHVHLQVKEPTRIIRQPEYKEVQRNSIAVFECKVKHDP
                        TLIPSMIWLKDNGELPDDPRFEVGSDSLTIHDVTEDDEGTYTCIRNTTLDQDSASATL
                        TVVEQPDDPPTDLELTDQRERSVRLTWTPGDEHNSPIKLFLIQYEDSLHEPGVWLNMTE
                        VSGTSTTAQLELSPYVYYSFRVLALNGVGLSESSDPSRQYRTNPAKPDVNPSDVEVSG
                        TSPDTMTISWRELSGLESNGPGLQYKVSWRMKDAEQWTTVTLANVSQHVVTGTPTFTL
                        YEVTVQAVNDYGEGPRPEVVLGYSGENSPTVAPENVKVSVQSGTEAEVHWEAVPLSSV
                        RGRLKGYKVTYQKMRSLHKQDYDRENPQVLIFSGEETVGRLPDLHPYSHYRLNIRAFN
                        GHGDGPSSTDQQFQTPEGVPGPPTNVNIRNLNLDSLLVEWTPPLEDNGHLTGYLLKYQ
                        PINTTEEAGLLKEVLLPANETSYTLDKLTHSTHYKFFLNAMTETGSGPAVTKEAFTEV
                        DEALIRHPAVEAGKAPPAGPMFGNVNSSVKEDHAVISWEYLGPDGNVYVEYVVDNSKE
                        PWKTEFVNGTRTFQIRGLKPGMSYRVRLVAKDHSDATIHSTQEMLITVPAMTNRQAEI
                        ATQGWFIGLMCAIALLILVLLIVCFIKRNKGGKYPVKEKEDAHQDPEIQPMKEDDGTF
                        GEYSDTEDHKPLKGSRTPSNGTVKKDDSDDSLVDYGEGGDGQFNEDGSFIGQYSGKKE
                        KDTAEGNESSEAPSPVNAMNSFV"
     gene               <82542. .>96655
                        /locus_tag="DKEY-240A12.2-001"
     mRNA               join(<82542. .82666,84647. .84828,84890. .85009,86543. .86749,
                        86849. .86948,87194. .87375,89617. .89741,89846. .89963,
                        90071. .90221,90302. .90429,90531. .90703,90784. .90907,
                        92537. .92704,93404. .93553,93627. .93697,93782. .94013,
                        95506. .95618,95698. .95887,95972. .96087,96497. .>96655)
                        /locus_tag="DKEY-240A12.2-001"
                        /product="novel similar to contactin 1 (cntn1)"
                        /note="match: ESTs: Em:CK027463.1"
     CDS                join(<82542. .82666,84647. .84828,84890. .85009,86543. .86749,
                        86849. .86948,87194. .87375,89617. .89741,89846. .89963,
                        90071. .90221,90302. .90429,90531. .90703,90784. .90907,
                        92537. .92704,93404. .93553,93627. .93697,93782. .94013,
                        95506. .95618,95698. .95887,95972. .96087,96497. .>96655)
                        /locus_tag="DKEY-240A12.2-001"
                        /standard_name="OTTDARP00000008715"
                        /note="match: proteins: Sw:P12960 Sw:P14781 Sw:P22063
                        Sw:P28685 Sw:Q12860 Sw:Q63198 Tr:O93250 Tr:P79921
                        Tr:Q8AXZ4"
                        /codon_start=1
                        /product="novel similar to contactin 1 (cntn1)"
                        /protein_id="CAH69054.1"
                        /db_xref="GI:55251199"
                        /db_xref="InterPro:IPR001827"
                        /db_xref="InterPro:IPR003598"
                        /db_xref="InterPro:IPR003599"
                        /db_xref="InterPro:IPR003961"
                        /db_xref="InterPro:IPR005479"
                        /db_xref="InterPro:IPR007110"
                        /db_xref="UniProt/TrEMBL:Q5TYT0"
                        /translation="EATGFGPIFEEEPLDTVYAEDSPENKISMNCRVRANPPAIIKWW
                        LNNWQIKLMEQPDEHFSLVGGNLVITNPEKNKHAGKYVCVAKNVYGTVMSKEANVKFG
                        FLVVLCSADLNQFPTDEREPVNVKEGQGAVLLCTPPSRYPDEVMFRWIYNDFPNFILP
                        DQRRFVSQSTGNLYIAKVEPSDVGNYSCFVSSPTIGKSVFSMPIALIPQIEKEVKRYP
                        ADIRVKFPKTYALVHQNITLECFALGNPLPHIRWRKLDADLPPNYEVSMNGALLHLIN
                        VQYEDEGSYECETLNVKGMDWYRQWLYVEGPPEWAEHINDTEKDVGSELTLRCVAVGK
                        PIPWIRWLKDGYSYGKGELKFSSLTFEDSGMYQCIAENEWGTIYANAELRVVSCAPTF
                        IYNPVKKILLGAENGRVVIECNPRAAPKPRFIWKRGSELLSNSSRIFIWDDGSLEILN
                        ATKSDEGSYTCYAENDRGKSNSTGTLTITEATKITVGPSNSEAAVGDTIVLQCSASYD
                        PSLDVTFIWTVDSYIINFYTDFEHYELDMDHENSGDLLIKNIQIRHAGHYICTAQTIV
                        DNATAEADVFVRGLPGPPGGVRVEEIGDTSVKIRWCLGSDHGSPLIQHIVQTRDFYAL
                        DPEDWKTATTTPAFLNGSTESATVVDLYPWMQYEFRVHAINEFGAGENSRPSIKIKTW
                        DARPTVAPSDVIGQVGISGELIVTWNPVKPQFFFGKKFGYVVAFKQHEDYEWKWSTVE
                        DPETRRYVYKESMTPDTEIQIKVSTFNNKGEGPASLISVVYSPRIAPTEAPLDVYARQ
                        VTSTEALVWWLPVYQAPPMWVDGYQIRYWRKYDDNEAAASRVIVHRTVNQTRLENMRP
                        DSHYLIEVRAFNGAGLGPPSEHCEMPTRRPPPSRRLRVYKYVSFTRKWLYLYWDHIYN
                        YWNESYVEGYKILFRKEGERYGKLYTTGRHYIDFPMPETGDYVIELRARCEGGDGPIS
                        QIRVQG"
```

```
     gene            153171. .>201932
                     /locus_tag="DKEY-240A12.3-001"
     mRNA            join(153171. .153240,186857. .187101,189855. .189957,
                     195148. .195246,197589. .197651,199444. .199545,
                     201816. .>201932)
                     /locus_tag="DKEY-240A12.3-001"
                     /product="novel protein"
                     /note="match: ESTs: Em:AL916334.1"
     CDS             join(153172. .153240,186857. .187101,189855. .189957,
                     195148. .195246,197589. .197651,199444. .199545,
                     201816. .>201932)
                     /locus_tag="DKEY-240A12.3-001"
                     /standard_name="OTTDARP00000008716"
                     /note="match: proteins: BAD16631 BAD18688 Tr:Q8NB75
                     Tr:Q9QY55 Tr:Q9UPQ7"
                     /codon_start=1
                     /product="novel protein"
                     /protein_id="CAH69055.1"
                     /db_xref="GI:55251200"
                     /db_xref="GOA:Q5TYS9"
                     /db_xref="InterPro:IPR001478"
                     /db_xref="UniProt/TrEMBL:Q5TYS9"
                     /translation="MGCNLCTLQKREEHYKLLYEIAQVNGREFSKSGHDGTVEALRKE
                     PIVVQVLRRGPSSVTQSLQEVCVVDVCTQTDITFEHIMALAKMRPSTPPVPDICPFLL
                     SDSCHSIQTMEHEYFDCPEYVASTQADVDRAVEFECEEVELCRMNSQEKLGLTLCYRT
                     DDEEDTAIFVGQVELNSIAARDGRIREGDRILQINGRDVQNREEAVALLSNENSRSIV

Alignment Scores:
Pred. No.:              3.58          Length:        201932
Score:                  84.00         Matches:       12
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     5             Gaps:          0

US-09-858-332G-15 (1-12) x BX897729 (1-201932)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     112864 CACAATCACAATCACAATCACAATCACAATCACAAT 112899

RESULT 83
AC069007/c
LOCUS       AC069007              204940 bp    DNA     linear   HTG 20-AUG-2000
DEFINITION  Homo sapiens chromosome 17 clone RP11-798G7, WORKING DRAFT
            SEQUENCE, 21 unordered pieces.
ACCESSION   AC069007
VERSION     AC069007.3  GI:9799889
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 204940)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 204940)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Aug 13, 2000 this sequence version replaced gi:9309540.

            ---------------- Genome Center ----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0798G07
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 191694 bases at least Q40
            Consensus quality: 196032 bases at least Q30
            Consensus quality: 198164 bases at least Q20
            Insert size: 229000; agarose-fp
            Insert size: 202940; sum-of-contigs
            Quality coverage: 4.53 in Q20 bases; agarose-fp
            Quality coverage: 5.19 in Q20 bases; sum-of-contigs
            ----------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 21 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1     1986: contig of 1986 bp in length
            *      1987     2086: gap of unknown length
            *      2087     3198: contig of 1112 bp in length
            *      3199     3298: gap of unknown length
            *      3299     5291: contig of 1993 bp in length
            *      5292     5391: gap of unknown length
            *      5392     7758: contig of 2367 bp in length
            *      7759     7858: gap of unknown length
            *      7859    12745: contig of 4887 bp in length
            *     12746    12845: gap of unknown length
            *     12846    17422: contig of 4577 bp in length
            *     17423    17522: gap of unknown length
            *     17523    22120: contig of 4598 bp in length
            *     22121    22220: gap of unknown length
            *     22221    26522: contig of 4302 bp in length
            *     26523    26622: gap of unknown length
            *     26623    31071: contig of 4449 bp in length
            *     31072    31171: gap of unknown length
            *     31172    36603: contig of 5432 bp in length
            *     36604    36703: gap of unknown length
            *     36704    42685: contig of 5982 bp in length
            *     42686    42785: gap of unknown length
            *     42786    49733: contig of 6948 bp in length
            *     49734    49833: gap of unknown length
            *     49834    60691: contig of 10858 bp in length
            *     60692    60791: gap of unknown length
            *     60792    73418: contig of 12627 bp in length
            *     73419    73518: gap of unknown length
            *     73519    89658: contig of 16140 bp in length
            *     89659    89758: gap of unknown length
            *     89759   102517: contig of 12759 bp in length
            *    102518   102617: gap of unknown length
            *    102618   121119: contig of 18502 bp in length
            *    121120   121219: gap of unknown length
            *    121220   126002: contig of 4783 bp in length
            *    126003   126102: gap of unknown length
            *    126103   146606: contig of 20504 bp in length
            *    146607   146706: gap of unknown length
            *    146707   171358: contig of 24652 bp in length
            *    171359   171458: gap of unknown length
            *    171459   204940: contig of 33482 bp in length.
FEATURES             Location/Qualifiers
     source          1. .204940
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /clone="RP11-798G7"
     misc_feature    1. .1986
                     /note="assembly_name:Contig10"
     gap             1987. .2086
                     /estimated_length=unknown
```

```
misc_feature    2087. .3198
                /note="assembly_name:Contig11"
gap             3199. .3298
                /estimated_length=unknown
misc_feature    3299. .5291
                /note="assembly_name:Contig12"
gap             5292. .5391
                /estimated_length=unknown
misc_feature    5392. .7758
                /note="assembly_name:Contig13"
gap             7759. .7858
                /estimated_length=unknown
misc_feature    7859. .12745
                /note="assembly_name:Contig14
                clone_end:T7
                vector_side:right"
gap             12746. .12845
                /estimated_length=unknown
misc_feature    12846. .17422
                /note="assembly_name:Contig15"
gap             17423. .17522
                /estimated_length=unknown
misc_feature    17523. .22120
                /note="assembly_name:Contig16"
gap             22121. .22220
                /estimated_length=unknown
misc_feature    22221. .26522
                /note="assembly_name:Contig17"
gap             26523. .26622
                /estimated_length=unknown
misc_feature    26623. .31071
                /note="assembly_name:Contig18"
gap             31072. .31171
                /estimated_length=unknown
misc_feature    31172. .36603
                /note="assembly_name:Contig19"
gap             36604. .36703
                /estimated_length=unknown
misc_feature    36704. .42685
                /note="assembly_name:Contig20"
gap             42686. .42785
                /estimated_length=unknown
misc_feature    42786. .49733
                /note="assembly_name:Contig21"
gap             49734. .49833
                /estimated_length=unknown
misc_feature    49834. .60691
                /note="assembly_name:Contig22"
gap             60692. .60791
                /estimated_length=unknown
misc_feature    60792. .73418
                /note="assembly_name:Contig23"
gap             73419. .73518
                /estimated_length=unknown
misc_feature    73519. .89658
                /note="assembly_name:Contig24"
gap             89659. .89758
                /estimated_length=unknown
misc_feature    89759. .102517
                /note="assembly_name:Contig25"
gap             102518. .102617
                /estimated_length=unknown
misc_feature    102618. .121119
                /note="assembly_name:Contig26"
gap             121120. .121219
                /estimated_length=unknown
misc_feature    121220. .126002
                /note="assembly_name:Contig27"
gap             126003. .126102
                /estimated_length=unknown
misc_feature    126103. .146606
                /note="assembly_name:Contig28"
gap             146607. .146706
                /estimated_length=unknown
misc_feature    146707. .171358
                /note="assembly_name:Contig29
                clone_end:SP6
                vector_side:right"
gap             171359. .171458
                /estimated_length=unknown
misc_feature    171459. .204940
                /note="assembly_name:Contig30"
ORIGIN

Alignment Scores:
Pred. No.:                3.63          Length:          204940
Score:                    84.00         Matches:         12
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       14            Gaps:            0

US-09-858-332G-15 (1-12) x AC069007 (1-204940)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     116458 CATAATCATAATCATAATCATAATCATAATCATAAT 116423

RESULT 84
AC091132/c
LOCUS       AC091132               207611 bp    DNA     linear   PRI 16-FEB-2003
DEFINITION  Homo sapiens chromosome 17, clone RP11-798G7, complete sequence.
ACCESSION   AC091132
VERSION     AC091132.16  GI:28395005
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 207611)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-798G7
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 207611)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
            Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K.,
            Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
            Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-APR-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 207611)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
```

```
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JAN-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 207611)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 16, 2003 this sequence version replaced gi:26292022.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L11996
                Center clone name: 798_G_7
            --------------.
FEATURES             Location/Qualifiers
     source          1. .207611
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-798G7"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   complement(14. .136)
                     /rpt_family="L1MC/D"
     repeat_region   complement(213. .537)
                     /rpt_family="AluSx"
     repeat_region   794. .872
                     /rpt_family="MIR3"
     repeat_region   complement(873. .1078)
                     /rpt_family="MER58A"
     repeat_region   1079. .1113
                     /rpt_family="MIR3"
     repeat_region   1191. .1492
                     /rpt_family="AluSx"
     repeat_region   complement(1664. .1948)
                     /rpt_family="AluSx"
     repeat_region   2095. .2116
                     /rpt_family="AT_rich"
     repeat_region   complement(2432. .2726)
                     /rpt_family="L1"
     repeat_region   complement(2730. .3017)
                     /rpt_family="AluSq"
     repeat_region   complement(3069. .3431)
                     /rpt_family="L1ME1"
     repeat_region   complement(3432. .3735)
                     /rpt_family="AluY"
     repeat_region   complement(3736. .3901)
                     /rpt_family="L1ME1"
     repeat_region   complement(3902. .4200)
                     /rpt_family="AluJo"
     repeat_region   complement(4201. .4279)
                     /rpt_family="L1ME1"
     repeat_region   complement(4280. .4573)
                     /rpt_family="AluY"
     repeat_region   complement(4576. .4881)
                     /rpt_family="AluJb"
     repeat_region   complement(4882. .5105)
                     /rpt_family="L1ME1"
     repeat_region   complement(5114. .5308)
                     /rpt_family="MER58B"
     repeat_region   complement(5446. .5967)
                     /rpt_family="L1MA4"
     repeat_region   complement(5974. .6104)
                     /rpt_family="FLAM_C"
     repeat_region   complement(6114. .6415)
                     /rpt_family="AluSx"
     repeat_region   complement(6439. .7049)
                     /rpt_family="L1ME1"
     repeat_region   7050. .7350
                     /rpt_family="AluSx"
     repeat_region   7359. .7654
                     /rpt_family="AluSx"
     repeat_region   complement(7655. .7672)
                     /rpt_family="L1ME1"
     repeat_region   complement(7797. .7913)
                     /rpt_family="FLAM_C"
     repeat_region   complement(7951. .8309)
                     /rpt_family="L1ME1"
     repeat_region   8310. .8338
                     /rpt_family="(TG)n"
     repeat_region   8339. .8486
                     /rpt_family="(TA)n"
     repeat_region   complement(8664. .8948)
                     /rpt_family="AluSc"
     repeat_region   8952. .9026
                     /rpt_family="L2"
     repeat_region   9245. .9599
                     /rpt_family="L1MDa"
     repeat_region   9623. .9738
                     /rpt_family="MIR"
     repeat_region   9739. .10031
                     /rpt_family="AluSx"
     repeat_region   complement(10406. .10447)
                     /rpt_family="L2"
     repeat_region   complement(11055. .11188)
                     /rpt_family="MER5A"
     repeat_region   11528. .11592
                     /rpt_family="MER5A"
     repeat_region   11593. .11748
                     /rpt_family="AluSq"
     repeat_region   complement(11761. .11875)
                     /rpt_family="FLAM_A"
     repeat_region   complement(11888. .12171)
                     /rpt_family="AluSx"
     repeat_region   complement(13010. .13302)
                     /rpt_family="AluSq"
     repeat_region   13423. .13443
```

```
                     /rpt_family="(TTTTTG)n"
     repeat_region   13633. .13929
                     /rpt_family="AluY"
     repeat_region   complement(14020. .14174)
                     /rpt_family="L2"
     repeat_region   15545. .15618
                     /rpt_family="(TTAA)n"
     repeat_region   complement(15619. .15901)

Alignment Scores:
Pred. No.:                3.68          Length:          207611
Score:                    84.00         Matches:         12
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       8             Gaps:            0

US-09-858-332G-15 (1-12) x AC091132 (1-207611)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     172803 CATAATCATAATCATAATCATAATCATAATCATAAT 172768

RESULT 85
AC120797/c
LOCUS       AC120797               210631 bp    DNA     linear   ROD 12-JUL-2005
DEFINITION  Mus musculus chromosome 18, clone RP24-81D18, complete sequence.
ACCESSION   AC120797
VERSION     AC120797.9  GI:70721520
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 210631)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 18, clone RP24-81D18
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 210631)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 210631)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-FEB-2004) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 210631)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-2004) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   5  (bases 1 to 210631)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
            Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J.,
            Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
            DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
            Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,
            Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,
            Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
            MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
            Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
            Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUL-2005) Broad Institute of MIT and Harvard, 320
            Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 12, 2005 this sequence version replaced gi:45544747.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Broad Institute of MIT and Harvard
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
```

```
            Contact: sequence_submissions@broad.mit.edu
          -------------- Project Information
            Center project name: L24984
            Center clone name: 81_D_18
          --------------.
FEATURES             Location/Qualifiers
     source          1. .210631
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="18"
                     /map="18"
                     /clone="RP24-81D18"
                     /clone_lib="RPCI-24 Male Mouse BAC"
     repeat_region   20. .106
                     /rpt_family="L1"
     unsure          complement(39. .43)
                     /note="<30 qual SNGL region"
     repeat_region   107. .129
                     /rpt_family="(TTAA)n"
     repeat_region   130. .1207
                     /rpt_family="L1"
     repeat_region   1206. .1257
                     /rpt_family="L1MB8"
     repeat_region   1264. .1365
                     /rpt_family="(GGGA)n"
     unsure          1284. .1336
                     /note="single clone coverage"
     repeat_region   1517. .1734
                     /rpt_family="L1MB8"
     repeat_region   complement(1793. .2164)
                     /rpt_family="MTD"
     repeat_region   2472. .2650
                     /rpt_family="B2_Mm2"
     repeat_region   2651. .2842
                     /rpt_family="B2_Mm2"
     repeat_region   3160. .3204
                     /rpt_family="(TAAAA)n"
     repeat_region   complement(4192. .4438)
                     /rpt_family="MER102b"
     repeat_region   complement(5055. .5240)
                     /rpt_family="B4"
     repeat_region   5710. .5772
                     /rpt_family="(TCC)n"
     repeat_region   5878. .6024
                     /rpt_family="B1_MM"
     repeat_region   complement(6159. .6245)
                     /rpt_family="MER58B"
     repeat_region   7018. .7060
                     /rpt_family="(TG)n"
     repeat_region   7060. .7094
                     /rpt_family="(GA)n"
     repeat_region   complement(7107. .7208)
                     /rpt_family="ID_B1"
     repeat_region   7271. .7468
                     /rpt_family="B3"
     repeat_region   complement(7500. .7628)
                     /rpt_family="BLACKJACK"
     repeat_region   7650. .7699
                     /rpt_family="(TTG)n"
     repeat_region   complement(7704. .7808)
                     /rpt_family="ID_B1"
     repeat_region   8011. .8114
                     /rpt_family="(GTTTG)n"
     repeat_region   complement(8557. .8730)
                     /rpt_family="L1ME"
     repeat_region   complement(8876. .9355)
                     /rpt_family="BLACKJACK"
     repeat_region   10244. .10524
                     /rpt_family="L1_MM"
     unsure          10426. .10505
                     /note="single clone coverage"
     repeat_region   10605. .10806
                     /rpt_family="B3"
     repeat_region   11680. .11713
                     /rpt_family="(TAAA)n"
     repeat_region   12090. .12194
                     /rpt_family="L2"
     repeat_region   12396. .12435

Alignment Scores:
Pred. No.:              3.73          Length:         210631
Score:                  84.00         Matches:        12
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     9             Gaps:           0

US-09-858-332G-15 (1-12) x AC120797 (1-210631)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db       7696 CACAACCACAACCACAACCACAACCACAACCACAAC 7661

RESULT 86
AC074051
LOCUS       AC074051               218133 bp    DNA     linear   PRI 03-APR-2003
DEFINITION  Homo sapiens chromosome 16 clone RP11-382N13, complete sequence.
ACCESSION   AC074051
VERSION     AC074051.4  GI:29501848
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 218133)
  AUTHORS   DOE Joint Genome Institute, Stanford Human Genome Center and Los
            Alamos National Laboratory.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 218133)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 218133)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-FEB-2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 218133)
  AUTHORS   DOE Joint Genome Institute, Stanford Human Genome Center and Los
            Alamos National Laboratory.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2003) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Apr 3, 2003 this sequence version replaced gi:18481992.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center and Los Alamos
            National Laboratory
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 100% of Sequence;
            Estimated Total Number of Errors is 0.
FEATURES             Location/Qualifiers
     source          1. .218133
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-382N13"
ORIGIN

Alignment Scores:
Pred. No.:              3.85          Length:         218133
```

```
Score:                  84.00       Matches:       12
Percent Similarity:     100.00%     Conservative:  0
Best Local Similarity:  100.00%     Mismatches:    0
Query Match:            100.00%     Indels:        0
DB:                     8           Gaps:          0

US-09-858-332G-15 (1-12) x AC074051 (1-218133)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     194179 CATAATCATAATCATAATCATAATCATAATCATAAT 194214

RESULT 87
AC084068
LOCUS       AC084068               220006 bp    DNA     linear   HTG 12-OCT-2000
DEFINITION  Mus musculus chromosome 16 clone RP23-99N9, WORKING DRAFT SEQUENCE,
            34 unordered pieces.
ACCESSION   AC084068
VERSION     AC084068.1  GI:10799417
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 220006)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Mouse
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 220006)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-OCT-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 0
            Center clone name: RPCI-23_99N9
            -------
            Summary Statistics
            Consensus quality: 163231 bases at least Q40
            Consensus quality: 186532 bases at least Q30
            Consensus quality: 197772 bases at least Q20
            Estimated insert size: 263300; agarose-fp estimation
            Estimated insert size: 216706; sum-of-contigs estimation
            Quality coverage:  3.06 in Q20 bases; agarose-fp estimation
            Quality coverage:  3.72 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 34 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1077: contig of 1077 bp in length
            *     1078     1177: gap of unknown length
            *     1178     2615: contig of 1438 bp in length
            *     2616     2715: gap of unknown length
            *     2716     3830: contig of 1115 bp in length
            *     3831     3930: gap of unknown length
            *     3931     5081: contig of 1151 bp in length
            *     5082     5181: gap of unknown length
            *     5182     6344: contig of 1163 bp in length
            *     6345     6444: gap of unknown length
            *     6445     8247: contig of 1803 bp in length
            *     8248     8347: gap of unknown length
            *     8348     9654: contig of 1307 bp in length
            *     9655     9754: gap of unknown length
            *     9755    12249: contig of 2495 bp in length
            *    12250    12349: gap of unknown length
            *    12350    14102: contig of 1753 bp in length
            *    14103    14202: gap of unknown length
            *    14203    16868: contig of 2666 bp in length
            *    16869    16968: gap of unknown length
            *    16969    18735: contig of 1767 bp in length
            *    18736    18835: gap of unknown length
            *    18836    21665: contig of 2830 bp in length
            *    21666    21765: gap of unknown length
            *    21766    23688: contig of 1923 bp in length
            *    23689    23788: gap of unknown length
            *    23789    25406: contig of 1618 bp in length
            *    25407    25506: gap of unknown length
            *    25507    28291: contig of 2785 bp in length
            *    28292    28391: gap of unknown length
            *    28392    31605: contig of 3214 bp in length
            *    31606 .  31705: gap of unknown length
            *    31706    36110: contig of 4405 bp in length
            *    36111    36210: gap of unknown length
            *    36211    38773: contig of 2563 bp in length
            *    38774    38873: gap of unknown length
            *    38874    43927: contig of 5054 bp in length
            *    43928    44027: gap of unknown length
            *    44028    48086: contig of 4059 bp in length
            *    48087    48186: gap of unknown length
            *    48187    52042: contig of 3856 bp in length
            *    52043    52142: gap of unknown length
            *    52143    56265: contig of 4123 bp in length
            *    56266    56365: gap of unknown length
            *    56366    61449: contig of 5084 bp in length
            *    61450    61549: gap of unknown length
            *    61550    72368: contig of 10819 bp in length
            *    72369    72468: gap of unknown length
            *    72469    79329: contig of 6861 bp in length
            *    79330    79429: gap of unknown length
            *    79430    85112: contig of 5683 bp in length
            *    85113    85212: gap of unknown length
            *    85213    91971: contig of 6759 bp in length
            *    91972    92071: gap of unknown length
            *    92072   102675: contig of 10604 bp in length
            *   102676   102775: gap of unknown length
            *   102776   112543: contig of 9768 bp in length
            *   112544   112643: gap of unknown length
            *   112644   130524: contig of 17881 bp in length
            *   130525   130624: gap of unknown length
            *   130625   143680: contig of 13056 bp in length
            *   143681   143780: gap of unknown length
            *   143781   169227: contig of 25447 bp in length
            *   169228   169327: gap of unknown length
            *   169328   193440: contig of 24113 bp in length
            *   193441   193540: gap of unknown length
            *   193541   220006: contig of 26466 bp in length.
FEATURES             Location/Qualifiers
     source          1. .220006
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="16"
                     /clone="RP23-99N9"
                     /clone_lib="RPCI mouse BAC library 23"
     gap             1078. .1177
                     /estimated_length=unknown
     gap             2616. .2715
                     /estimated_length=unknown
     gap             3831. .3930
                     /estimated_length=unknown
     gap             5082. .5181
                     /estimated_length=unknown
     gap             6345. .6444
                     /estimated_length=unknown
     gap             8248. .8347
                     /estimated_length=unknown
```

```
     gap             9655. .9754
                     /estimated_length=unknown
     gap             12250. .12349
                     /estimated_length=unknown
     gap             14103. .14202
                     /estimated_length=unknown
     gap             16869. .16968
                     /estimated_length=unknown
     gap             18736. .18835
                     /estimated_length=unknown
     gap             21666. .21765
                     /estimated_length=unknown
     gap             23689. .23788
                     /estimated_length=unknown
     gap             25407. .25506
                     /estimated_length=unknown
     gap             28292. .28391
                     /estimated_length=unknown
     gap             31606. .31705
                     /estimated_length=unknown
     gap             36111. .36210
                     /estimated_length=unknown
     gap             38774. .38873
                     /estimated_length=unknown
     gap             43928. .44027
                     /estimated_length=unknown
     gap             48087. .48186
                     /estimated_length=unknown
     gap             52043. .52142
                     /estimated_length=unknown
     gap             56266. .56365
                     /estimated_length=unknown
     gap             61450. .61549
                     /estimated_length=unknown
     gap             72369. .72468
                     /estimated_length=unknown
     gap             79330. .79429
                     /estimated_length=unknown
     gap             85113. .85212
                     /estimated_length=unknown
     gap             91972. .92071
                     /estimated_length=unknown
     gap             102676. .102775
                     /estimated_length=unknown
     gap             112544. .112643
                     /estimated_length=unknown
     gap             130525. .130624
                     /estimated_length=unknown
     gap             143681. .143780
                     /estimated_length=unknown
     gap             169228. .169327
                     /estimated_length=unknown
     gap             193441. .193540
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              3.89          Length:         220006
Score:                  84.00         Matches:        12
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     14            Gaps:           0

US-09-858-332G-15 (1-12) x AC084068 (1-220006)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      47332 CATAACCATAACCATAACCATAACCATAACCATAAC 47367

RESULT 88
AC073680/c
LOCUS       AC073680              226013 bp    DNA     linear   HTG 29-JUN-2000
DEFINITION  Mus musculus clone RP23-116M1, WORKING DRAFT SEQUENCE, 20 unordered
            pieces.
ACCESSION   AC073680
VERSION     AC073680.1  GI:8810297
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 226013)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Mouse
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 226013)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 1781705
            Center clone name: RPCI-23_116M1
            -------
            Summary Statistics
            Consensus quality: 209118 bases at least Q40
            Consensus quality: 215883 bases at least Q30
            Consensus quality: 217616 bases at least Q20
            Estimated insert size: 216000; agarose-fp estimation
            Estimated insert size: 224113; sum-of-contigs estimation
            Quality coverage:  9.65 in Q20 bases; agarose-fp estimation
            Quality coverage:  9.3 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 20 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1157: contig of 1157 bp in length
            *     1158     1257: gap of unknown length
            *     1258     2308: contig of 1051 bp in length
            *     2309     2408: gap of unknown length
            *     2409     3408: contig of 1000 bp in length
            *     3409     3508: gap of unknown length
            *     3509     4765: contig of 1257 bp in length
            *     4766     4865: gap of unknown length
            *     4866     6182: contig of 1317 bp in length
            *     6183     6282: gap of unknown length
            *     6283     7697: contig of 1415 bp in length
            *     7698     7797: gap of unknown length
            *     7798     9195: contig of 1398 bp in length
            *     9196     9295: gap of unknown length
            *     9296    10579: contig of 1284 bp in length
            *    10580    10679: gap of unknown length
            *    10680    12073: contig of 1394 bp in length
            *    12074    12173: gap of unknown length
            *    12174    13684: contig of 1511 bp in length
            *    13685    13784: gap of unknown length
            *    13785    14932: contig of 1148 bp in length
            *    14933    15032: gap of unknown length
            *    15033    17640: contig of 2608 bp in length
            *    17641    17740: gap of unknown length
            *    17741    19342: contig of 1602 bp in length
            *    19343    19442: gap of unknown length
            *    19443    21366: contig of 1924 bp in length
            *    21367    21466: gap of unknown length
            *    21467    27246: contig of 5780 bp in length
```

```
            *    27247    27346: gap of unknown length
            *    27347    30988: contig of 3642 bp in length
            *    30989    31088: gap of unknown length
            *    31089    35750: contig of 4662 bp in length
            *    35751    35850: gap of unknown length
            *    35851    50689: contig of 14839 bp in length
            *    50690    50789: gap of unknown length
            *    50790    70274: contig of 19485 bp in length
            *    70275    70374: gap of unknown length
            *    70375   226013: contig of 155639 bp in length.
FEATURES             Location/Qualifiers
     source          1. .226013
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-116M1"
                     /clone_lib="RPCI mouse BAC library 23"
     gap             1158. .1257
                     /estimated_length=unknown
     gap             2309. .2408
                     /estimated_length=unknown
     gap             3409. .3508
                     /estimated_length=unknown
     gap             4766. .4865
                     /estimated_length=unknown
     gap             6183. .6282
                     /estimated_length=unknown
     gap             7698. .7797
                     /estimated_length=unknown
     gap             9196. .9295
                     /estimated_length=unknown
     gap             10580. .10679
                     /estimated_length=unknown
     gap             12074. .12173
                     /estimated_length=unknown
     gap             13685. .13784
                     /estimated_length=unknown
     gap             14933. .15032
                     /estimated_length=unknown
     gap             17641. .17740
                     /estimated_length=unknown
     gap             19343. .19442
                     /estimated_length=unknown
     gap             21367. .21466
                     /estimated_length=unknown
     gap             27247. .27346
                     /estimated_length=unknown
     gap             30989. .31088
                     /estimated_length=unknown
     gap             35751. .35850
                     /estimated_length=unknown
     gap             50690. .50789
                     /estimated_length=unknown
     gap             70275. .70374
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              3.99          Length:         226013
Score:                  84.00         Matches:        12
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     14            Gaps:           0

US-09-858-332G-15 (1-12) x AC073680 (1-226013)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      96991 CACAACCACAACCACAACCACAACCACAACCACAAC 96956

RESULT 89
AC119847/c
```

```
LOCUS       AC119847               227081 bp    DNA     linear   ROD 27-SEP-2003
DEFINITION  Mus musculus chromosome 3, clone RP23-46L5, complete sequence.
ACCESSION   AC119847
VERSION     AC119847.12  GI:34330229
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 227081)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 3, clone RP23-46L5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 227081)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-MAY-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 227081)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-AUG-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 227081)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   5  (bases 1 to 227081)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,X., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-2003) Whitehead Institute/MIT Center for Genome
```

```
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Aug 29, 2003 this sequence version replaced gi:33590182.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L22334
                Center clone name: 46_L_5
            --------------.
FEATURES             Location/Qualifiers
     source          1. .227081
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="3"
                     /map="3"
                     /clone="RP23-46L5"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1. .15469
                     /note="wgs_end_extension
                     clone_end:SP6"
     repeat_region   1. .275
                     /rpt_family="L1MC4a"
     repeat_region   696. .786
                     /rpt_family="PB1D9"
     repeat_region   789. .824
                     /rpt_family="(A)n"
     repeat_region   complement(896. .1056)
                     /rpt_family="B3"
     repeat_region   1194. .1219
                     /rpt_family="(CA)n"
     repeat_region   complement(1381. .1492)
                     /rpt_family="B3"
     repeat_region   1493. .1906
                     /rpt_family="MTC"
     repeat_region   complement(1907. .2009)
                     /rpt_family="B3"
     repeat_region   complement(2016. .2156)
                     /rpt_family="B1F"
     repeat_region   2165. .2198
                     /rpt_family="AT_rich"
     repeat_region   2476. .2564
                     /rpt_family="(TG)n"
     repeat_region   2681. .3076
                     /rpt_family="MTA"
     repeat_region   3144. .3228
                     /rpt_family="ID_B1"
     repeat_region   complement(3649. .3782)
                     /rpt_family="B1_MM"
     repeat_region   4619. .4728
                     /rpt_family="PB1"
     repeat_region   4784. .4907
                     /rpt_family="B1_MM"
     repeat_region   5040. .5060
                     /rpt_family="AT_rich"
     repeat_region   5064. .5221
                     /rpt_family="B2_Mm2"
     repeat_region   5222. .5248
                     /rpt_family="(TAAA)n"
     repeat_region   5249. .5275
                     /rpt_family="B2_Mm2"
     repeat_region   6264. .6307
                     /rpt_family="AT_rich"
     repeat_region   6686. .6916
                     /rpt_family="MTE"
     repeat_region   complement(7237. .7370)
                     /rpt_family="B2_Mm2"
     repeat_region   7612. .7763
                     /rpt_family="B2_Mm1"
     repeat_region   7764. .7790
                     /rpt_family="(TAAA)n"
     repeat_region   7791. .7818
                     /rpt_family="B2_Mm1"
     repeat_region   8132. .8225
                     /rpt_family="PB1D9"
     repeat_region   8236. .8302
                     /rpt_family="RSINE1"
     repeat_region   complement(8302. .8472)
                     /rpt_family="ORR1D"
     repeat_region   8482. .8536
                     /rpt_family="(TC)n"
     repeat_region   8537. .8574
                     /rpt_family="(CA)n"
     repeat_region   8894. .9040
                     /rpt_family="ORR1D"
     repeat_region   complement(9105. .9238)
                     /rpt_family="ID_B1"
     repeat_region   9267. .9460
                     /rpt_family="B3"
     repeat_region   9601. .9643
                     /rpt_family="(TTA)n"
     repeat_region   9848. .10053
                     /rpt_family="B4A"
     repeat_region   10054. .10237
                     /rpt_family="B2_Mm2"
     repeat_region   10238. .10292
                     /rpt_family="B4A"
     repeat_region   10307. .10352
                     /rpt_family="(CA)n"
     repeat_region   10540. .10702
                     /rpt_family="RSINE1"
     repeat_region   complement(10957. .10970)
                     /rpt_family="B4"
     repeat_region   complement(10971. .11052)
                     /rpt_family="PB1D9"
     repeat_region   complement(11053. .11090)
                     /rpt_family="B4"
     repeat_region   complement(11065. .11188)
                     /rpt_family="ID_B1"
     repeat_region   complement(11194. .11267)
                     /rpt_family="ID5"
     repeat_region   11383. .11406
                     /rpt_family="PB1D10"
     repeat_region   11407. .11494
                     /rpt_family="ID_B1"
     repeat_region   complement(11925. .12030)
                     /rpt_family="B1F"
     repeat_region   complement(12265. .12445)
                     /rpt_family="B3"
     repeat_region   complement(12536. .12731)
                     /rpt_family="B3"
     unsure          complement(12878. .12880)

Alignment Scores:
Pred. No.:              4.01          Length:         227081
Score:                  84.00         Matches:        12
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     9             Gaps:           0

US-09-858-332G-15 (1-12) x AC119847 (1-227081)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     192501 CATAACCATAACCATAACCATAACCATAACCATAAC 192466

RESULT 90
AC133743
LOCUS       AC133743              227648 bp    DNA     linear   HTG 20-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-182C23, *** SEQUENCING IN PROGRESS
```

```
            ***, 9 unordered pieces.
ACCESSION   AC133743
VERSION     AC133743.2  GI:25139560
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 227648)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,X.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 227648)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 227648)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 20, 2002 this sequence version replaced gi:23096694.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: KAPR
                Center clone name: CH230-182C23
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 189235 bases at least Q40
                Consensus quality: 193444 bases at least Q30
                Consensus quality: 195848 bases at least Q20
                Estimated insert size: 190572; sum-of-contigs estimation
                Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    50224: contig of 50224 bp in length
            *    50225    50324: gap of unknown length
            *    50325    53089: contig of 2765 bp in length
            *    53090    53189: gap of unknown length
            *    53190   128906: contig of 75717 bp in length
            *   128907   129006: gap of unknown length
            *   129007   220174: contig of 91168 bp in length
            *   220175   220274: gap of unknown length
            *   220275   221473: contig of 1199 bp in length
            *   221474   221573: gap of unknown length
            *   221574   223286: contig of 1713 bp in length
            *   223287   223386: gap of unknown length
            *   223387   224587: contig of 1201 bp in length
            *   224588   224687: gap of unknown length
            *   224688   225923: contig of 1236 bp in length
            *   225924   226023: gap of unknown length
            *   226024   227648: contig of 1625 bp in length.
FEATURES             Location/Qualifiers
     source          1. .227648
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-182C23"
     gap             50225. .50324
                     /estimated_length=unknown
     gap             53090. .53189
                     /estimated_length=unknown
     gap             128907. .129006
                     /estimated_length=unknown
     misc_feature    129007. .130011
                     /note="wgs_contig"
     misc_feature    165940. .167729
                     /note="wgs_contig"
     misc_feature    198830. .199863
```

```
                /note="wgs_contig"
     misc_feature    203077. .204817
                /note="wgs_contig"
     gap             220175. .220274
                /estimated_length=unknown
     gap             221474. .221573
                /estimated_length=unknown
     gap             223287. .223386
                /estimated_length=unknown
     gap             224588. .224687
                /estimated_length=unknown
     gap             225924. .226023
                /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              4.02          Length:         227648
Score:                  84.00         Matches:        12
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     14            Gaps:           0

US-09-858-332G-15 (1-12) x AC133743 (1-227648)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      98443 CACAACCACAACCACAACCACAACCACAACCACAAC 98478

RESULT 91
AC154121
LOCUS       AC154121              228353 bp    DNA     linear   ROD 31-MAR-2005
DEFINITION  Mus musculus chromosome 1, clone RP24-93A17, complete sequence.
ACCESSION   AC154121
VERSION     AC154121.6  GI:62079390
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 228353)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 1, clone RP24-93A17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 228353)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
            Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J.,
            Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
            DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
            Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,
            Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,
            Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
            MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
            Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
            Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-DEC-2004) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 228353)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
            Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J.,
            Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
            DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
            Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,
            Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,
            Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
            MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
            Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
            Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-2005) Broad Institute of MIT and Harvard, 320
            Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 31, 2005 this sequence version replaced gi:61696474.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Broad Institute of MIT and Harvard
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@broad.mit.edu
            -------------- Project Information
                Center project name: L30878
                Center clone name: 93_A_17
            --------------.
FEATURES             Location/Qualifiers
     source          1. .228353
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="1"
                     /map="1"
                     /clone="RP24-93A17"
                     /clone_lib="RPCI-24 Male Mouse BAC"
     repeat_region   complement(32. .171)
                     /rpt_family="B1_MM"
     repeat_region   174. .201
                     /rpt_family="AT_rich"
```

```
repeat_region   324. .601
                /rpt_family="B4A"
repeat_region   complement(893. .1143)
                /rpt_family="MTC"
repeat_region   1148. .1205
                /rpt_family="CT-rich"
repeat_region   1213. .1280
                /rpt_family="(TCC)n"
repeat_region   complement(1282. .1403)
                /rpt_family="ORR1D"
repeat_region   1498. .1656
                /rpt_family="(GAAA)n"
repeat_region   complement(2435. .2543)
                /rpt_family="RMER1B"
repeat_region   2556. .2703
                /rpt_family="B1_MM"
repeat_region   complement(2746. .2967)
                /rpt_family="RMER1B"
repeat_region   complement(2997. .3218)
                /rpt_family="RMER1B"
repeat_region   complement(3241. .3326)
                /rpt_family="MTC"
repeat_region   3327. .3358
                /rpt_family="(A)n"
repeat_region   4043. .4068
                /rpt_family="AT_rich"
repeat_region   5640. .6566
                /rpt_family="RMER17B"
repeat_region   6585. .6622
                /rpt_family="(CA)n"
repeat_region   6765. .6793
                /rpt_family="AT_rich"
repeat_region   7495. .7517
                /rpt_family="AT_rich"
repeat_region   8090. .8126
                /rpt_family="AT_rich"
repeat_region   complement(8582. .8671)
                /rpt_family="ORR1D"
repeat_region   8884. .8974
                /rpt_family="(CA)n"
repeat_region   9525. .9930
                /rpt_family="Lx4"
repeat_region   10137. .10254
                /rpt_family="L1M2"
repeat_region   10255. .10294
                /rpt_family="(TAAA)n"
repeat_region   complement(10854. .11213)
                /rpt_family="MTE"
repeat_region   11639. .11811
                /rpt_family="MIR"
repeat_region   complement(12388. .14373)
                /rpt_family="L1"
repeat_region   14367. .15182
                /rpt_family="L1_MM"
repeat_region   15689. .17956
                /rpt_family="Lx4"
repeat_region   complement(17958. .18599)
                /rpt_family="L1_RN"
repeat_region   18600. .18626
                /rpt_family="(TTA)n"
repeat_region   complement(18627. .18771)
                /rpt_family="B1_MM"
repeat_region   complement(18772. .19900)
                /rpt_family="L1_RN"
repeat_region   19938. .19968
                /rpt_family="(CATA)n"
repeat_region   22022. .22076
                /rpt_family="(CA)n"
repeat_region   22090. .22213
                /rpt_family="Lx9"
repeat_region   22944. .22999
                /rpt_family="CT-rich"
repeat_region   23017. .23183
                /rpt_family="(TG)n"
repeat_region   23428. .23456
                /rpt_family="AT_rich"
repeat_region   23513. .23555
                /rpt_family="(TC)n"
repeat_region   23559. .23602
                /rpt_family="(CA)n"
repeat_region   25189. .25248
                /rpt_family="(TG)n"
repeat_region   25961. .26006
                /rpt_family="AT_rich"
repeat_region   26153. .26191

Alignment Scores:
Pred. No.:                4.03          Length:          228353
Score:                    84.00         Matches:         12
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       9             Gaps:            0

US-09-858-332G-15 (1-12) x AC154121 (1-228353)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     190914 CATAATCATAATCATAATCATAACCATAATCATAAT 190949

RESULT 92
CT009527
LOCUS       CT009527              228485 bp    DNA     linear   HTG 10-AUG-2005
DEFINITION  Mus musculus chromosome 13 clone RP23-25D11, WORKING DRAFT
            SEQUENCE, 3 unordered pieces.
ACCESSION   CT009527
VERSION     CT009527.2  GI:72416650
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 228485)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 12, 2005 this sequence version replaced gi:68226311.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            -------------- Project Information
            Center project name: bM25D11
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 228251 bases at least Q40
            Consensus quality: 228267 bases at least Q30
            Consensus quality: 228276 bases at least Q20
            Insert size: 228285; sum-of-contigs
            Insert size: 202389; 5.5% error; agarose-fp
            Quality coverage: 11.01x in Q20 bases; sum-of-contigs Quality
            coverage: 12.41x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
```

```
            *       1   114901: contig of 114901 bp in length
            *  .114902   115001: gap of 100 bp
            *   115002   178742: contig of 63741 bp in length
            *   178743   178843: gap of 101 bp
            *   178844   228485: contig of 49642 bp in length.
FEATURES             Location/Qualifiers
     source          1. .228485
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="13"
                     /clone="RP23-25D11"
                     /clone_lib="RPCI-23"
     misc_feature    1. .114901
                     /note="assembly_fragment:01528
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    115002. .178742
                     /note="assembly_fragment:00610
                     fragment_chain:1"
     misc_feature    178843. .228485
                     /note="assembly_fragment:00007
                     fragment_chain:1
                     clone_end:T7
                     vector_side:right"
ORIGIN

Alignment Scores:
Pred. No.:                4.03          Length:         228485
Score:                    84.00         Matches:        12
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              100.00%       Indels:         0
DB:                       14            Gaps:           0

US-09-858-332G-15 (1-12) x CT009527 (1-228485)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     120068 CACAACCACAACCACAACCACAACCACAACCACAAC 120103

RESULT 93
AC096973/c
LOCUS       AC096973               231383 bp    DNA     linear   HTG 13-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-2N7, WORKING DRAFT SEQUENCE, 3
            unordered pieces.
ACCESSION   AC096973
VERSION     AC096973.7  GI:24942103
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 231383)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 231383)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 231383)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 13, 2002 this sequence version replaced gi:23683156.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GAWU
                Center clone name: CH230-2N7
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 216032 bases at least Q40
                Consensus quality: 217638 bases at least Q30
                Consensus quality: 218825 bases at least Q20
                Estimated insert size: 223315; sum-of-contigs estimation
```

```
            Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            -------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   215694: contig of 215694 bp in length
            *  215695   215794: gap of unknown length
            *  215795   218857: contig of 3063 bp in length
            *  218858   218957: gap of unknown length
            *  218958   231383: contig of 12426 bp in length.
FEATURES             Location/Qualifiers
     source          1. .231383
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-2N7"
     misc_feature    1. .1209
                     /note="wgs_contig"
     misc_feature    complement(206387. .206798)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:
                     end_sequence:BH289712"
     misc_feature    211671. .215694
                     /note="wgs_end_extension
                     clone_end:Sp6"
     gap             215695. .215794
                     /estimated_length=unknown
     gap             218858. .218957
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:                4.08          Length:          231383
Score:                    84.00         Matches:         12
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       14            Gaps:            0

US-09-858-332G-15 (1-12) x AC096973 (1-231383)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      47404 CACAACCACAACCACAACCACAACCACAACCACAAC 47369

RESULT 94
AC159263
LOCUS       AC159263               231384 bp    DNA     linear   HTG 16-JUN-2005
DEFINITION  Mus musculus chromosome 13 clone RP24-66K21, WORKING DRAFT
            SEQUENCE, 11 unordered pieces.
ACCESSION   AC159263
VERSION     AC159263.2  GI:67515043
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 231384)
  AUTHORS   Wilson,R.K.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 231384)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-2005) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 231384)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUN-2005) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Jun 13, 2005 this sequence version replaced gi:62079489.

            ---------------- Genome Center ----------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu
            Contact: submissions@watson.wustl.edu
            -------------- Project Information ------------------
            Center project name: M_BB0066K21
            ----------------------------------------------------
            ------------- Summary Statistics  -------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 224638 bases at least Q40
            Consensus quality: 225270 bases at least Q30
            Consensus quality: 225719 bases at least Q20
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1037: contig of 1037 bp in length
            *    1038     1137: gap of unknown length
            *    1138     3356: contig of 2219 bp in length
            *    3357     3456: gap of unknown length
            *    3457    11997: contig of 8541 bp in length
            *   11998    12097: gap of unknown length
            *   12098    16930: contig of 4833 bp in length
            *   16931    17030: gap of unknown length
            *   17031    31608: contig of 14578 bp in length
            *   31609    31708: gap of unknown length
            *   31709    55615: contig of 23907 bp in length
            *   55616    55715: gap of unknown length
            *   55716    78818: contig of 23103 bp in length
            *   78819    78918: gap of unknown length
            *   78919   108247: contig of 29329 bp in length
            *  108248   108347: gap of unknown length
            *  108348   143432: contig of 35085 bp in length
            *  143433   143532: gap of unknown length
            *  143533   181352: contig of 37820 bp in length
            *  181353   181452: gap of unknown length
            *  181453   231384: contig of 49932 bp in length.
FEATURES             Location/Qualifiers
     source          1. .231384
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="13"
                     /clone="RP24-66K21"
     misc_feature    1. .1037
                     /note="assembly_name:Contig20"
     gap             1038. .1137
                     /estimated_length=unknown
     misc_feature    1138. .3356
                     /note="assembly_name:Contig38"
     gap             3357. .3456
                     /estimated_length=unknown
     misc_feature    3457. .11997
                     /note="assembly_name:Contig40"
```

```
     gap             11998. .12097
                     /estimated_length=unknown
     misc_feature    12098. .16930
                     /note="assembly_name:Contig41"
     gap             16931. .17030
                     /estimated_length=unknown
     misc_feature    17031. .31608
                     /note="assembly_name:Contig42"
     gap             31609. .31708
                     /estimated_length=unknown
     misc_feature    31709. .55615
                     /note="assembly_name:Contig43"
     gap             55616. .55715
                     /estimated_length=unknown
     misc_feature    55716. .78818
                     /note="assembly_name:Contig44"
     gap             78819. .78918
                     /estimated_length=unknown
     misc_feature    78919. .108247
                     /note="assembly_name:Contig45
                     clone_end:T7
                     vector_side:right"
     gap             108248. .108347
                     /estimated_length=unknown
     misc_feature    108348. .143432
                     /note="assembly_name:Contig46"
     gap             143433. .143532
                     /estimated_length=unknown
     misc_feature    143533. .181352
                     /note="assembly_name:Contig47"
     gap             181353. .181452
                     /estimated_length=unknown
     misc_feature    181453. .231384
                     /note="assembly_name:Contig48"
ORIGIN

Alignment Scores:
Pred. No.:                4.08          Length:          231384
Score:                    84.00         Matches:         12
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       14            Gaps:            0

US-09-858-332G-15 (1-12) x AC159263 (1-231384)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     123917 CATAACCATAACCATAACCATAACCATAACCATAAC 123952

RESULT 95
AC117096
LOCUS       AC117096               232413 bp    DNA     linear   HTG 19-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-268L9, *** SEQUENCING IN PROGRESS
            ***, 7 unordered pieces.
ACCESSION   AC117096
VERSION     AC117096.4  GI:25072937
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 232413)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 232413)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 232413)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 19, 2002 this sequence version replaced gi:23194579.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
```

```
            Center project name: GTKW
            Center clone name: CH230-268L9
          ---------- Summary Statistics
            Assembly program: Phrap; version 0.990329
            Consensus quality: 212301 bases at least Q40
            Consensus quality: 214075 bases at least Q30
            Consensus quality: 215384 bases at least Q20
            Estimated insert size: 215514; sum-of-contigs estimation
            Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
          ----------
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 7 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1    42591: contig of 42591 bp in length
          *    42592    42691: gap of unknown length
          *    42692   189134: contig of 146443 bp in length
          *   189135   189234: gap of unknown length
          *   189235   199851: contig of 10617 bp in length
          *   199852   199951: gap of unknown length
          *   199952   222653: contig of 22702 bp in length
          *   222654   222753: gap of unknown length
          *   222754   227954: contig of 5201 bp in length
          *   227955   228054: gap of unknown length
          *   228055   229970: contig of 1916 bp in length
          *   229971   230070: gap of unknown length
          *   230071   232413: contig of 2343 bp in length.
FEATURES             Location/Qualifiers
     source          1. .232413
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-268L9"
     misc_feature    1. .1737
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    complement(4465. .5309)
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:BZ173680"
     gap             42592. .42691
                     /estimated_length=unknown
     gap             189135. .189234
                     /estimated_length=unknown
     misc_feature    189235. .191036
                     /note="wgs_contig"
     misc_feature    197729. .198806
                     /note="wgs_contig"
     gap             199852. .199951
                     /estimated_length=unknown
     misc_feature    199952. .201761
                     /note="wgs_contig"
     gap             222654. .222753
                     /estimated_length=unknown
     misc_feature    222754. .224182
                     /note="wgs_contig"
     misc_feature    224233. .227954
                     /note="wgs_contig"
     gap             227955. .228054
                     /estimated_length=unknown
     gap             229971. .230070
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              4.1             Length:         232413
```

```
Score:                  84.00           Matches:        12
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-15 (1-12) x AC117096 (1-232413)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     228695 CACAACCACAACCACAACCACAACCACAACCACAAC 228730

RESULT 96
CR450845
LOCUS       CR450845              235546 bp    DNA     linear   HTG 29-JUL-2005
DEFINITION  Danio rerio chromosome 2 clone DKEY-63K7, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION   CR450845
VERSION     CR450845.11  GI:71533119
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 235546)
  AUTHORS   Hunter,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUL-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests:
            http://www.sanger.ac.uk/Projects/D_rerio/faqs.shtml#dataeight
COMMENT     On Jul 29, 2005 this sequence version replaced gi:70671332.
            ---------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            ---------- Project Information
            Center project name: zK63K7
            ---------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 235360 bases at least Q40
            Consensus quality: 235405 bases at least Q30
            Consensus quality: 235426 bases at least Q20
            Insert size: 235446; sum-of-contigs
            Insert size: 232920; 2.9% error; agarose-fp
            Quality coverage: 9.20x in Q20 bases; sum-of-contigs Quality
            coverage: 9.30x in Q20 bases; agarose-fp
            ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1   152972: contig of 152972 bp in length
            *   152973   153072: gap of 100 bp
            *   153073   235546: contig of 82474 bp in length.
FEATURES             Location/Qualifiers
     source          1. .235546
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /chromosome="2"
                     /clone="DKEY-63K7"
                     /clone_lib="DanioKey"
     misc_feature    1. .152972
                     /note="assembly_fragment:03594
                     fragment_chain:1"
```

```
     misc_feature    153073. .235546
                     /note="assembly_fragment:00328
                     fragment_chain:1"
ORIGIN

Alignment Scores:
Pred. No.:              4.15          Length:        235546
Score:                  84.00         Matches:       12
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     14            Gaps:          0

US-09-858-332G-15 (1-12) x CR450845 (1-235546)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     127426 CATAATCATAATCATAATCATAATCATAATCATAAT 127461

RESULT 97
AC127135/c
LOCUS       AC127135               237732 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-89C20, *** SEQUENCING IN PROGRESS
            ***, 3 unordered pieces.
ACCESSION   AC127135
VERSION     AC127135.4  GI:30520824
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 237732)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 237732)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 237732)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:24819332.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GHZF
                Center clone name: CH230-89C20
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 206671 bases at least Q40
                Consensus quality: 211484 bases at least Q30
                Consensus quality: 214331 bases at least Q20
                Estimated insert size: 220281; sum-of-contigs estimation
                Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1   234751: contig of 234751 bp in length
            *    234752   234851: gap of unknown length
            *    234852   235859: contig of 1008 bp in length
            *    235860   235959: gap of unknown length
            *    235960   237732: contig of 1773 bp in length.
FEATURES             Location/Qualifiers
     source          1. .237732
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-89C20"
```

```
     misc_feature    1. .1165
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    complement(7252. .8117)
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH337952"
     misc_feature    135628. .138318
                     /note="wgs_contig"
     misc_feature    164949. .167852
                     /note="wgs_contig"
     misc_feature    229390. .230234
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH337953"
     misc_feature    230944. .234751
                     /note="wgs_end_extension
                     clone_end:Sp6"
     gap             234752. .234851
                     /estimated_length=unknown
     gap             235860. .235959
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:                4.19          Length:          237732
Score:                    84.00         Matches:         12
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       14            Gaps:            0

US-09-858-332G-15 (1-12) x AC127135 (1-237732)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     198279 CACAACCACAACCACAACCACAACCACAACCACAAC 198244

RESULT 98
AC097541
LOCUS       AC097541               240663 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-27J1, *** SEQUENCING IN PROGRESS ***,
            5 unordered pieces.
ACCESSION   AC097541
VERSION     AC097541.7  GI:30522471
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 240663)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 240663)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 240663)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:24956660.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GGQT
                Center clone name: CH230-27J1
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 215135 bases at least Q40
                Consensus quality: 220155 bases at least Q30
                Consensus quality: 223518 bases at least Q20
                Estimated insert size: 228408; sum-of-contigs estimation
                Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
            --------------
```

```
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1   232805: contig of 232805 bp in length
            *   232806   232905: gap of unknown length
            *   232906   234010: contig of 1105 bp in length
            *   234011   234110: gap of unknown length
            *   234111   235542: contig of 1432 bp in length
            *   235543   235642: gap of unknown length
            *   235643   237928: contig of 2286 bp in length
            *   237929   238028: gap of unknown length
            *   238029   240663: contig of 2635 bp in length.
FEATURES             Location/Qualifiers
     source          1. .240663
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-27J1"
     misc_feature    1. .2549
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    4456. .5233
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH337345"
     misc_feature    4496. .74792
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH337344"
     misc_feature    60177. .61683
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    108980. .111715
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    227776. .232805
                     /note="wgs_end_extension
                     clone_end:T7"
     gap             232806. .232905
                     /estimated_length=unknown
     gap             234011. .234110
                     /estimated_length=unknown
     gap             235543. .235642
                     /estimated_length=unknown
     gap             237929. .238028
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              4.24          Length:         240663
Score:                  84.00         Matches:        12
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     14            Gaps:           0

US-09-858-332G-15 (1-12) x AC097541 (1-240663)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      27493 CACAACCACAACCACAACCACAACCACAACCACAAC 27528

RESULT 99
AC106512/c

LOCUS       AC106512              240773 bp    DNA     linear   HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-201G13, WORKING DRAFT SEQUENCE.
ACCESSION   AC106512
VERSION     AC106512.6  GI:30580799
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 240773)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 240773)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 240773)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
```

```
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:23265140.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            ---------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            ---------- Project Information
                Center project name: GLBJ
                Center clone name: CH230-201G13
            ---------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 229749 bases at least Q40
                Consensus quality: 231863 bases at least Q30
                Consensus quality: 233224 bases at least Q20
                Estimated insert size: 241516; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            ----------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   240773: contig of 240773 bp in length.
FEATURES            Location/Qualifiers
     source         1. .240773
                    /organism="Rattus norvegicus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10116"
                    /clone="CH230-201G13"
     misc_feature   1. .2580
                    /note="wgs_end_extension
                    clone_end:T7"
     misc_feature   3883. .7533
                    /note="wgs_end_extension
                    clone_end:T7"
     misc_feature   complement(8181. .8589)
                    /note="clone_boundary
                    clone_end:T7
                    site:EcoRI
                    end_sequence:BH331454"
     misc_feature   238394. .239041
                    /note="clone_boundary
                    clone_end:Sp6
                    site:EcoRI
                    end_sequence:BH331455"
ORIGIN

Alignment Scores:
Pred. No.:              4.24          Length:         240773
Score:                  84.00         Matches:        12
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     14            Gaps:           0

US-09-858-332G-15 (1-12) x AC106512 (1-240773)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     238012 CACAACCACAACCACAACCACAACCACAACCACAAC 237977

RESULT 100
AC152062/c
LOCUS       AC152062              242626 bp    DNA     linear   HTG 21-APR-2005
DEFINITION  Mus musculus chromosome 10 clone RP23-341K9, WORKING DRAFT
            SEQUENCE, 3 unordered pieces.
ACCESSION   AC152062
VERSION     AC152062.5  GI:62822109
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 242626)
  AUTHORS   Wilson,R.K.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 242626)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2004) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 242626)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-APR-2005) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Apr 21, 2005 this sequence version replaced gi:59958552.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu
            Contact: submissions@watson.wustl.edu
            ---------- Project Information ----------
            Center project name: M_BA0341K09
            ----------
            ---------- Summary Statistics  ----------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 239082 bases at least Q40
            Consensus quality: 240311 bases at least Q30
            Consensus quality: 242181 bases at least Q20
            ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1503: contig of 1503 bp in length
            *    1504     1603: gap of unknown length
            *    1604    76669: contig of 75066 bp in length
            *   76670    76769: gap of unknown length
            *   76770   242626: contig of 165857 bp in length.
FEATURES            Location/Qualifiers
     source         1. .242626
                    /organism="Mus musculus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10090"
                    /chromosome="10"
```

```
                     /clone="RP23-341K9"
     misc_feature    1. .1503
                     /note="assembly_name:Contig18"
     gap             1504. .1603
                     /estimated_length=unknown
     misc_feature    1604. .76669
                     /note="assembly_name:Contig26"
     gap             76670. .76769
                     /estimated_length=unknown
     misc_feature    76770. .242626
                     /note="assembly_name:Contig27"
ORIGIN

Alignment Scores:
Pred. No.:              4.27          Length:         242626
Score:                  84.00         Matches:        12
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     14            Gaps:           0

US-09-858-332G-15 (1-12) x AC152062 (1-242626)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     103653 CACAACCACAACCACAACCACAACCACAACCACAAC 103618

RESULT 101
AY491413
LOCUS       AY491413              243772 bp    DNA     linear   HTG 23-FEB-2004
DEFINITION  Mus musculus chromosome 17 clone BAC 229p12, *** SEQUENCING IN
            PROGRESS ***, 4 ordered pieces.
ACCESSION   AY491413
VERSION     AY491413.1  GI:42602054
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 243772)
  AUTHORS   Brathwaite,M., Waeltz,P., Dudekula,D. and Nagaraja,R.
  TITLE     Genomic Sequence Analysis in the Mouse t-complex Region
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 243772)
  AUTHORS   Brathwaite,M., Waeltz,P., Dudekula,D. and Nagaraja,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-NOV-2003) Laboratory of Genetics, National Institute
            on Aging/NIH, 333 Cassell Drive, Suite 3000, Baltimore, MD 21224,
            USA
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    53217: contig of 53217 bp in length
            *    53218    53317: gap of 100 bp
            *    53318    61652: contig of 8335 bp in length
            *    61653    61752: gap of 100 bp
            *    61753   123635: contig of 61883 bp in length
            *   123636   123735: gap of 100 bp
            *   123736   243772: contig of 120037 bp in length.
FEATURES             Location/Qualifiers
     source          1. .243772
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="17"
                     /clone="BAC 229p12"
                     /clone_lib="CITB CJ7"
     repeat_region   351. .513
                     /rpt_family="SINE/B4"
     repeat_region   549. .686
                     /rpt_family="LINE/L1"
     repeat_region   complement(1196. .1283)
                     /rpt_family="LINE/L1"
     repeat_region   complement(1331. .1583)
                     /rpt_family="SINE/B4"
     repeat_region   complement(1626. .1834)
                     /rpt_family="SINE/B2"
     repeat_region   complement(1903. .2040)
                     /rpt_family="SINE/Alu"
     repeat_region   complement(2080. .2201)
                     /rpt_family="SINE/Alu"
     repeat_region   complement(2518. .2653)
                     /rpt_family="SINE/B4"
     repeat_region   complement(2656. .2767)
                     /rpt_family="SINE/B2"
     repeat_region   2897. .3026
                     /rpt_family="SINE/Alu"
     repeat_region   complement(5432. .5497)
                     /rpt_family="SINE/Alu"
     repeat_region   complement(8439. .8621)
                     /rpt_family="DNA/MER1_type"
     repeat_region   complement(11278. .11419)
                     /rpt_family="SINE/Alu"
     mRNA            join(<13693. .13870,14737. .14927,15393. .15488,15664. .15796,
                     16250. .16393,17492. .17620,17744. .17926,19455. .19661,
                     19934. .20146,20229. .20395,21802. .21943,25949. .26705)
                     /product="unknown"
                     /note="similar to Mus musculus mRNA GI:37359997"
                     /evidence=not_experimental
     CDS             join(<13693. .13870,14737. .14927,15393. .15488,15664. .15796,
                     16250. .16393,17492. .17620,17744. .17926,19455. .19661,
                     19934. .20146,20229. .20395,21802. .21943,25949. .26158)
                     /note="similar to Homo sapiens KIAA0590; similar to
                     GI:7662194; coding sequence predicted by Grailexp"
                     /codon_start=2
                     /evidence=not_experimental
                     /product="unknown"
                     /protein_id="AAS21643.1"
                     /db_xref="GI:42602055"
                     /translation="EAVWENMARMCVKTQRLDVAKVCLGNMGHARGARALREAEQEPE
                     LEARVAMLAIQLGMLEEAEQLYKKCKRYDLLNKFYQASDQWQKAVEVAELHDRVHLRT
                     TYYNYAKHLEASADCGQALSYYEKSDTHRFEVPRMLSEDLQSLELYINRMKDKTLWRW
                     WAQYLESQAEMDTALRYYELAQDYFSLVRIHCFQGNIQKAAEIANETGDWAASYHLAR
                     QYESQDEVKQAVHFYTRAQAFNNAIRLCKENSLDDQLMNLALLSSPEDMIEAARYYEE
                     KGEQMDRAVMLYHKAGHFSKALELAFTTQQFAALQLIAEDLDEKSDPALLARCSDFCI
                     EHRQFEKAVELLLAAKKYHEALQLCLEQNMTITEDMAEKMTVSKDSKDMSEESRRELL
                     EQIANCCMRQGNYHLATKKYTQAGNKLKAMRALLKSGDTEKIVFFAGVSRQKEIYIMA
                     ANYLQSLDWRKEPEIMKSIISFYTKGRALDLLAGFYDACAQVEIDEYQNYDKAHGALT
                     EAYKCLSKAKTKNPLDQETKLAQLQSKMTLVKRFIQARRTYTEDPKESLRQCELLLEE
                     PDLDSTIRVGDVYGFLVEHHVQMEEYQMAYKYLEEMRKRLPSANMSYYVDQRTVDTVH
                     QGLGLLPPSRIMPERVRHNSMEDHKEVYEEVIEEVDNDP"
     repeat_region   15962. .16161
                     /rpt_family="SINE/B2"
     repeat_region   16875. .17022
                     /rpt_family="SINE/Alu"
     repeat_region   17992. .18139
                     /rpt_family="SINE/Alu"
     repeat_region   complement(18871. .18980)
                     /rpt_family="SINE/Alu"
     repeat_region   complement(19023. .19158)
                     /rpt_family="SINE/Alu"
     repeat_region   complement(20660. .20991)
                     /rpt_family="LTR/MaLR"
     repeat_region   complement(22025. .22279)
                     /rpt_family="LTR/MaLR"
     repeat_region   22280. .22499
                     /rpt_family="SINE/B2"
     repeat_region   complement(22500. .22578)
                     /rpt_family="LTR/MaLR"
```

```
repeat_region   complement(22649. .22752)
                /rpt_family="SINE/B4"
repeat_region   complement(22766. .22835)
                /rpt_family="SINE/ID"
repeat_region   23815. .23956
                /rpt_family="SINE/Alu"
repeat_region   complement(24276. .24507)
                /rpt_family="LTR/ERVK"
repeat_region   complement(24502. .24601)
                /rpt_family="LTR/ERVK"
repeat_region   24602. .24744
                /rpt_family="SINE/Alu"
repeat_region   complement(24745. .25084)
                /rpt_family="LTR/ERVK"
repeat_region   25191. .25507
                /rpt_family="LINE/L1"
polyA_signal    26681. .26687
                /evidence=not_experimental
mRNA            complement(join(26787. .27338,28526. .28641,28867. .28931,
                29238. .29337,29864. .29955,30806. .30997,31803. .31874,
                32004. .32120,32280. .32371,32542. .32630,32926. .33039,
                33402. .33481,34191. .34327,34408. .34549,34746. .34814,
                37677. .37779,38157. .38304,38570. .38638,40446. .40723,
                42393. .42741,42869. .43184))
                /product="unknown"
                /note="similar to Riken cDNA 120000M09"
                /evidence=not_experimental
polyA_signal    complement(26808. .26813)
                /note="Grailexp prediction"
                /evidence=not_experimental
CDS             complement(join(27226. .27338,28526. .28641,28867. .28931,
                29238. .29337,29864. .29955,30806. .30997,31803. .31874,
                32004. .32120,32280. .32371,32542. .32630,32926. .33039,
                33402. .33481,34191. .34327,34408. .34549,34746. .34814,
                37677. .37779,38157. .38304,38570. .38638,40446. .40723,
                42393. .42727))
                /note="similar to Riken cDNA 1200003M09, GI:15029722,
                Grailexp  predicted coding sequence; 1200003M09Rik
                protein"
                /codon_start=1
                /evidence=not_experimental
                /product="unknown"
                /protein_id="AAS21644.1"
                /db_xref="GI:42602056"
                /translation="MDPALSAVRLTVQEAIHTLSSSEDAGHILSTLGTLKRYLGGTED
                PVLPEEKEEFATVHFSAVLRCLVSKLSPGWLELSPGGQLERLWESFFLDGPPDQAFLV
                LMEAIESTAGPSFRLMKMAQLLDTFLSTGRVAALMEEQCRPQTKPSFPLFQETLLSKV
                VGLPDLLGNCLQRDNLTQFFPQNYFPLLGQEVVQALKAVVNFLQGGLDCSVSFVSRVL
                GKVCIQGRKREILSVLVPQLTVLTQDSCLWQRVCWRLVEQVPDRAVEAVLTGLVEAAP
                RPEVLSRLLGNLVVKNKKARFVVTRKLLLLQYQHTTPMVQSLLGYLALDSQRRPLLIQ
                VLKELLETWGCSSAVRHTPLEQQCYISKAILVCLAHLGEPELQDIRDELLASMMAGVK
                CRLDSSLPPVRRLGMIVAEVISSRIHPEGPLLKFQYEDDEMSRELLALATPEPAGDCS
                SVSRGPSPAPADTEPPVEMPEKAVESDVPPTQPQGSDSELDSDDEFIPYDMSGDRELK
                NSKEPLYIRDCVEALTTSEDMERWEASLKGLEGLVYRSPAATREVSVELAKVLLHLEE
                KTCVAEFEQLRQSALVAVTVTDPEQVAKYLTSQFYGLNYSLRQRMDILDVLVLAAQAL
                SRPKSLQRRSQHGSLVPGTMCSPALAVSQTGNVAAPDWQMVVEERIRSKTRRFSKGCP
                QRELSGVPNEFSSVAGYFFFPLLQHFDRPLVTFDLLGDDQLVLGRLTHTLASLMYLAV
                NTTVAVPMGKALLEFVWALRFHVDIYVRRGLLSAVSSVLLSVPTERLLGDLPDELLEA
                RSWLADVAEKDVDEDCRELAVRALLLLERLKDKLLSSSSPQP"
repeat_region   complement(27922. .27976)
                /rpt_family="SINE/B2"
repeat_region   complement(27990. .28166)
                /rpt_family="SINE/B2"
repeat_region   complement(28172. .28308)
                /rpt_family="SINE/Alu"
repeat_region   complement(28207. .28331)
                /rpt_family="SINE/B4"
repeat_region   complement(30304. .30498)
                /rpt_family="SINE/B2"
repeat_region   complement(30631. .30789)
                /rpt_family="DNA/MER1_type"
repeat_region   30994. .31064
                /rpt_family="SINE/B2"
repeat_region   31063. .31156
                /rpt_family="SINE/B2"
repeat_region   31358. .31513
                /rpt_family="SINE/B2"
repeat_region   complement(33119. .33311)
                /rpt_family="SINE/B2"
repeat_region   complement(35172. .35336)
                /rpt_family="SINE/B2"
repeat_region   complement(35337. .35452)
                /rpt_family="SINE/Alu"
```

Alignment Scores:

Pred. No.:	4.29	Length:	243772
Score:	84.00	Matches:	12
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	14	Gaps:	0

US-09-858-332G-15 (1-12) x AY491413 (1-243772)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     243636 CACAACCACAACCACAACCACAACCACAACCACAAC 243671
```

```
RESULT 102
AC094502/c
LOCUS       AC094502               248752 bp    DNA     linear   HTG 09-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-4E8, *** SEQUENCING IN PROGRESS ***,
            3 unordered pieces.
ACCESSION   AC094502
VERSION     AC094502.6  GI:30466791
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 248752)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
```

Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 248752)

AUTHORS Worley,K.C.

TITLE Direct Submission

JOURNAL Submitted (15-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 248752)

AUTHORS Rat Genome Sequencing Consortium.

TITLE Direct Submission

JOURNAL Submitted (09-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On May 9, 2003 this sequence version replaced gi:24818812.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GATW
Center clone name: CH230-4E8
---------- Summary Statistics
Assembly program: Atlas;
Consensus quality: 206523 bases at least Q40
Consensus quality: 211412 bases at least Q30
Consensus quality: 214233 bases at least Q20
Estimated insert size: 215006; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 245823: contig of 245823 bp in length
* 245824 245923: gap of unknown length
* 245924 247299: contig of 1376 bp in length
* 247300 247399: gap of unknown length
* 247400 248752: contig of 1353 bp in length.

FEATURES Location/Qualifiers

source 1. .248752
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-4E8"

misc_feature 1. .1697
/note="wgs_end_extension
clone_end:T7"

misc_feature 2836. .3642
/note="clone_boundary
clone_end:T7
site:EcoRI
end_sequence:BH305316"

misc_feature 90194. .91759
/note="wgs_contig"

misc_feature 91810. .93031
/note="wgs_contig"

misc_feature 152801. .154271
/note="wgs_contig"

misc_feature 184165. .185672
/note="wgs_contig"

misc_feature complement(237880. .238728)
/note="clone_boundary
clone_end:Sp6
site:EcoRI
end_sequence:BH305317"

misc_feature 244828. .245823
/note="wgs_end_extension
clone_end:Sp6"

gap 245824. .245923
/estimated_length=unknown

gap 247300. .247399
/estimated_length=unknown

ORIGIN

Alignment Scores:

Pred. No.:	4.37	Length:	248752
Score:	84.00	Matches:	12
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	14	Gaps:	0

US-09-858-332G-15 (1-12) x AC094502 (1-248752)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     136235 CACAACCACAACCACAACCACAACCACAACCACAAC 136200
```

RESULT 103
AC113885/c
LOCUS AC113885 248784 bp DNA linear HTG 15-NOV-2002
DEFINITION Rattus norvegicus clone CH230-336B24, *** SEQUENCING IN PROGRESS ***, 2 unordered pieces.
ACCESSION AC113885
VERSION AC113885.5 GI:24635526
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 248784)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,

Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 248784)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (05-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 248784)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Nov 6, 2002 this sequence version replaced gi:23816056.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GSZX
Center clone name: CH230-336B24
---------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 205078 bases at least Q40
Consensus quality: 207520 bases at least Q30
Consensus quality: 208922 bases at least Q20
Estimated insert size: 212129; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 247293: contig of 247293 bp in length
* 247294 247393: gap of unknown length
* 247394 248784: contig of 1391 bp in length.

FEATURES Location/Qualifiers
source 1. .248784
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-336B24"
misc_feature 1. .1534
/note="wgs_end_extension
clone_end:Sp6"
misc_feature 41356. .42217
/note="clone_boundary
clone_end:Sp6
site:
end_sequence:BZ114916"
misc_feature 245604. .246444
/note="clone_boundary
clone_end:T7
site:
end_sequence:BZ114915"
gap 247294. .247393
/estimated_length=unknown
ORIGIN

Alignment Scores:

Pred. No.:	4.38	Length:	248784
Score:	84.00	Matches:	12
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	14	Gaps:	0

US-09-858-332G-15 (1-12) x AC113885 (1-248784)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     136466 CATAATCATAATCATAATCATAATCATAATCACAAC 136431
```

RESULT 104
AC097616/c
LOCUS AC097616 250591 bp DNA linear HTG 13-MAY-2003
DEFINITION Rattus norvegicus clone CH230-184D3, *** SEQUENCING IN PROGRESS ***, 2 unordered pieces.
ACCESSION AC097616
VERSION AC097616.6 GI:30581287
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 250591)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 250591)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (20-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 250591)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On May 13, 2003 this sequence version replaced gi:23268166.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
-------------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-------------- Project Information
Center project name: GISB
Center clone name: CH230-184D3
-------------- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 232007 bases at least Q40
Consensus quality: 235142 bases at least Q30
Consensus quality: 237284 bases at least Q20
Estimated insert size: 260427; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 34592: contig of 34592 bp in length
* 34593 34692: gap of unknown length
* 34693 250591: contig of 215899 bp in length.

FEATURES Location/Qualifiers
source 1. .250591
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-184D3"
misc_feature 1. .1654
/note="wgs_contig"
misc_feature 1705. .4775
/note="wgs_contig"
misc_feature 5962. .9772
/note="wgs_contig"
misc_feature 9823. .12235
/note="wgs_contig"
misc_feature 12286. .18055
/note="wgs_contig"
misc_feature 18106. .19463
/note="wgs_contig"
misc_feature 19682. .25998
/note="wgs_contig"
misc_feature 27570. .29058
/note="wgs_contig"
misc_feature 32458. .34592
/note="wgs_contig"
gap 34593. .34692
/estimated_length=unknown
misc_feature 34693. .36406
/note="wgs_contig"
ORIGIN

Alignment Scores:

Pred. No.:	4.41	Length:	250591
Score:	84.00	Matches:	12
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	14	Gaps:	0

US-09-858-332G-15 (1-12) x AC097616 (1-250591)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     117483 CATAACCACAACCACAACCACAACCACAACCACAAC 117448

RESULT 105
AC095012
LOCUS       AC095012               252619 bp    DNA     linear   HTG 09-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-6O11, WORKING DRAFT SEQUENCE.
ACCESSION   AC095012
VERSION     AC095012.6  GI:30467631
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 252619)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 252619)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 252619)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 9, 2003 this sequence version replaced gi:24818257.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GBYY
                Center clone name: CH230-6O11
            -------------- Summary Statistics
                Assembly program: Atlas;
                Consensus quality: 234170 bases at least Q40
                Consensus quality: 236579 bases at least Q30
                Consensus quality: 238129 bases at least Q20
                Estimated insert size: 244153; sum-of-contigs estimation
                Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *         1   252619: contig of 252619 bp in length.
FEATURES             Location/Qualifiers
     source          1. .252619
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-6O11"
     misc_feature    1. .1083
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    complement(8178. .8925)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH270975"
     misc_feature    complement(247639. .248424)
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH270974"
     misc_feature    251420. .252619
                     /note="wgs_end_extension
                     clone_end:T7"
ORIGIN
```

```
Alignment Scores:
Pred. No.:              4.44          Length:        252619
Score:                  84.00         Matches:       12
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     14            Gaps:          0

US-09-858-332G-15 (1-12) x AC095012 (1-252619)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     113685 CATAATCATAATCATAATCATAATCATAATCATAAT 113720

RESULT 106
AC126742
LOCUS       AC126742               253494 bp    DNA     linear   HTG 09-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-4G23, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION   AC126742
VERSION     AC126742.6  GI:30467824
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 253494)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 253494)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 253494)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 9, 2003 this sequence version replaced gi:24941150.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GCGW
                Center clone name: CH230-4G23
            -------------- Summary Statistics
                Assembly program: Atlas;
                Consensus quality: 212667 bases at least Q40
                Consensus quality: 215613 bases at least Q30
                Consensus quality: 217844 bases at least Q20
                Estimated insert size: 224651; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   252279: contig of 252279 bp in length
            *  252280   252379: gap of unknown length
            *  252380   253494: contig of 1115 bp in length.
FEATURES             Location/Qualifiers
     source          1. .253494
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-4G23"
     misc_feature    1. .1098
                     /note="wgs_contig"
     misc_feature    2071. .4080
                     /note="wgs_contig"
     misc_feature    4131. .5739
                     /note="wgs_contig"
     gap             252280. .252379
```

```
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:                4.45          Length:          253494
Score:                    84.00         Matches:         12
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       14            Gaps:            0

US-09-858-332G-15 (1-12) x AC126742 (1-253494)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     204196 CACAACCACAACCACAACCACAACCACAACCACAAC 204231

RESULT 107
AC117075
LOCUS       AC117075               254733 bp    DNA     linear   INV 12-MAR-2003
DEFINITION  Dictyostelium discoideum chromosome 2 map 5201047-5455781 strain
            AX4, complete sequence.
ACCESSION   AC117075 AC115615 AC117177
VERSION     AC117075.2  GI:28828046
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 254733)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 254733)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-APR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 254733)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 254733)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On or before Mar 4, 2003 this sequence version replaced
            gi:20069690, gi:19570144, gi:20066242.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES              Location/Qualifiers
     source           1. .254733
                      /organism="Dictyostelium discoideum"
                      /mol_type="genomic DNA"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /chromosome="2"
                      /map="5201047-5455781"
     CDS              complement(join(<124. .619,719. .1068,1167. .1457))
                      /note="GeneID exon scores (in order of location ranges):
                      30.56, 24.41, 16.56 - GSCJ_ID dd_01823"
                      /codon_start=1
                      /product="hypothetical protein"
                      /protein_id="AAM09355.2"
                      /db_xref="GI:28828135"
                      /translation="MGIEGLNLFMGKQFGCVEKVKEIPTDHVYIDLNNYLHKSVKRGD
                      NSKLNTEVNVFRILKSFIDGILRKVRVKHSVFFGIDGPGPRSKMILQRERRLKNGNID
                      KLKYFINRNKEQQQQQQQQQQSPQLHDYDYLNENDLKSFDSSFSTLEFTPGTTFMGK
                      LKDFLIFYTKNKLYFAKKIFISASDRIGEGEFKIFEQILNSNYPINDSFTIVSSDSDI
                      LLFSLLSKYKNVYIYNKDSEEIIKIDKIRDKIYQQCYKKKQKKQKQQKKSGGEVEGEG
                      EGIATEEKEEGVQIEEEEEEEEEEEEDISKRRQAIVDFVMLTFLMGTDHLPKVSSYNI
                      GSAWSEYCKIKKPLYNEETGLINLDLLFRLIGKSSQPRLFYRNRF"
     CDS              complement(join(1796. .3073,3249. .3563))
                      /note="GeneID exon scores (in order of location ranges):
                      87.30, 9.45 - GSCJ_ID dd_01824"
                      /codon_start=1
                      /product="hypothetical protein"
                      /protein_id="AAO50731.1"
                      /db_xref="GI:28828047"
                      /translation="MKLGRDIPIYKLIIIKNESIENNLKEISFSDNFKCQICEGLLIS
                      SLIPNRMKALQCINGHCFCLTCWESILEIKSECPTCRIQIQSMNTLSNNLFIIKSISE
                      SIKIHCPNYLNFDNSNNFNGCKEIITIDEIDRHESKCEFRFIKCSINNQCNEIIRFNE
                      RDKHESQCDFIIQKCTHCDESVQMKQMQGHILFECLKVLITCQHCSLQFTRLKLQNHI
                      ENHCKEIKIECPFKSLKIIYNNNNNDNDNNDSDENNSNQSLSSSSLSSSIIPLSPCNE
                      LMKRSELSKHFIRNHQYHNELVSIVLKKQENKIKKLETIIQQNNVNQNFEISIIKHSN
                      SLTRDSFYKENKKKDKLIQDLMKRVLILEHQQQQNQIQNQIQNQIQKLSNKSIFTTTT
                      TTTSTSDKNNNNNNNNNNNNNNNEDEDDDNIKNNDNQGNLKSLREMFNTISNFKDD
                      ENEQQQIQQQQQLPFTSFGPLSVQQQNNPTKQFNQLSQPQTQPQSQSQSQSLFGEWSP
                      ITTINQNQNTPSNPFSIFSGTSL"
     CDS              complement(join(4148. .4536,4589. .4883,4975. .5261,
                      5357. .5513,5597. .5830,5946. .6098))
                      /note="GeneID exon scores (in order of location ranges):
                      33.35, 9.09, 28.75, 10.40, 9.66, 11.24 - GSCJ_ID dd_01825"
                      /codon_start=1
                      /product="similar to Mus musculus (Mouse). 0 day neonate
                      eyeball cDNA, RIKEN full-length enriched library,
                      clone:E130103I15 product:polycystic kidney disease 2, full
                      insert sequence"
                      /protein_id="AAO50732.1"
                      /db_xref="GI:28828048"
                      /translation="MARSGYYIDIPVSDIKTGVESLIDNGFFDIQTRSVIISFSTLNL
                      NFQSRTSSGQHQDQLIPIISIRTYRIQLYMDALDKFRGFLEVSFFIFILYYFYYFVNE
                      ARIDFKLGRFKFYISNFKNIFDVINLTLFIASIALYLAFLGDNSRDLNINSQIQTTTD
                      YPLYLENLGQTALVLYQISAINILLMSFKTFRFLVVHRRLYILWIAISQSRLQLITFT
                      IMFCIMMLGFLFSGWLTFGTDIASFNGFVTSLGTLLQFIIGNPPDYEAMSYTNRALGP
                      IYYLLFTIFMFFILVNMFVAIISDSYQNISNTFEKKQKTKRYLLTRWERTMKSLLLLF
                      TPNSLDIYDLVLMFIDQKPGLIETDGIDPSRFKDEIMIIQPIQRARIKYLDQIQSAKN
                      TGVTFELKLMGQQVHKKDLDTWKKEAKIKEKKKFESIQQNINILNTKLDLLLQAMNVP
                      QQSIPQLINGSTGMPQSPIINNNNNNGGSIPSSSSNGSINGSNHINLSGGTNTP"
     CDS              complement(join(6215. .6478,6568. .6864))
                      /note="GeneID exon scores (in order of location ranges):
                      21.78, 25.12 - GSCJ_ID dd_03502"
                      /codon_start=1
                      /product="hypothetical protein"
                      /protein_id="AAO50733.1"
                      /db_xref="GI:28828049"
                      /translation="MNTLKRTVFSQRLEAVKALENKTREEKATLLAAVETQRNARLVK
                      DLIIHLVFFVVFVLIVVLQLNPQSLFIINTSLKNQILFSQDSQYLEINNPSDFITYLN
                      STFCESISDMSLENNDAMNKIIGNVIRIRTARVKPDSCSNNGFNLTCYSNNYDKDTKD
                      RSPFGPNDMYTYTSNSHGSYVFGMFF"
     CDS              join(8148. .8280,8374. .10196)
                      /note="GeneID exon scores (in order of location ranges):
                      7.51, 117.32 - GSCJ_ID dd_00414"
                      /codon_start=1
                      /product="similar to Suppressor of mutant AC40 subunit of
                      RNA polymerase I and III (high serine); Srp40p
                      [Saccharomyces cerevisiae]"
                      /protein_id="AAO50734.1"
                      /db_xref="GI:28828050"
                      /translation="MNKTLSNKKASSKTTNNSNNNNNNTINDPVLSLPTYSLEIYEED
                      YFNELILEPLSFLKSGDVVITTSPTKKKTSSVSASPSPIKKESNSNSNSNSNSNSN
                      SNNNNNDSDSSSDNDLKFEGNQDDSEDEDQTNSNYYCPFCKKKFQTEQTWNTHINSSK
                      HLKSIKDQKKKSTTTGTSNTPTKNNNNNSPSNVKSPSSTKTKSSSSSSSSSSSSSSSS
                      SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSTNSKDNEFLQDQIMQTQQAIKIKNSKP
```

NLACKNLFTIARSYCQKQYIRDSAKTLNLIIDILTISNNNNNNNTTPSITSTTTSSTS
NKSIVEYSLNNILNNATLLCKTHLYFGRLTRLFNQRLSEYHFLQCFNYLALINNDQLL
KYESSLVSSSPLNLYKIFIIQLDKFGKSKNKDTNTNTDTDTNTNTDTNKEIESIIQIN
KYINEIAGALCYCEVNFKSIILYILASCITFKFKLYKESLESNTRAFKLLSSLKQYHW
SIDLLLNNLLIKLKLNETIDIVDGDGDTEIQYQKLLLKEKEEKEEKEYQNIIIILIKY
CILIDDKIRLSAILGDYNFQKIISIYFKNNDYVDCLIKSSNQFLIENYKEIDQIKDHY
EHLIYSNKDFIFIFNQFDLKKKILLQK"

CDS complement(join(10448. .10915,11145. .11294))
/note="GeneID exon scores (in order of location ranges): 45.58, 25.53 - GSCJ_ID dd_00412"
/codon_start=1
/product="similar to Dictyostelium discoideum (Slime mold). Ras-related protein Rap-1"
/protein_id="AAO50735.1"
/db_xref="GI:28828051"
/translation="MGKGNGKSKDSKGSEKSSTKTKFFVAVMGAGSVGKSALTVQFTQ
GVFIDKYDPTVEDTYTKTFELDGEQVCIEVLDTAGSEVLVAMRELYMKNAEGFILVYS
ILVKSTFLEIKDIVEQLFRVKEEEEVPIVLVGNKIDLDAHREVSTNEGKQLANSYPNC
DFWEASSKDRINVDNVFFSIVKRIKEKYKKEGAPVKEKKEKCIIM"

CDS 12794. .13423
/note="GeneID exon scores (in order of location ranges): 35.19 - GSCJ_ID dd_00410"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAO50736.1"
/db_xref="GI:28828052"
/translation="MASIGTGKFKNEPTFGKNIKEILILLERSIAESEKQWVDLVSEN
KNLVDVNAFRINKLFDNKIGLDVKDNKSFELLKKTVHEFKTERPVIDLVCKSISCLFY
LSDLQLSSIITNLPNEIIAQLKDSTVPFAYRLENSSLKNVQIQKRHYLSSNRPIELVC
CILYLPANIDIICKVVPKSKFLNETSISGCPFDLEMDSSFYGIKKKLSL"

CDS complement(join(13494. .13813,13915. .14097,14208. .14241))
/note="GeneID exon scores (in order of location ranges): 22.37, 12.24, 8.90 - GSCJ_ID dd_00408"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAO50737.1"
/db_xref="GI:28828053"
/translation="MSIYEEIMRAQVKQGNMGGRSGPLGYLLRENKQKEKKAVKVNTP
KPPKPVVTQSPYTINLIEGTVTNATDNNGFSDPYLVIIAIKPDNSFSKILFKSNVCKK
TLTPTWNQVENFFVKDNVTDIIVELWDHDVLSKNDFIGRSIISMNTVRSKGYDGTLEF
FNGSEVVDGTVKLTIKRK"

CDS complement(join(14655. .14686,14848. .14916,14992. .15162, 15268. .15310))
/note="GeneID exon scores (in order of location ranges): 1.29, 3.37, 1.88, 4.01 - GSCJ_ID dd_00406"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAO50738.1"
/db_xref="GI:28828054"
/translation="MSEICSTSCWQYAIDNPNCFPKIGFTKVCFMSGNYTCASSSGFT
ILSDELSFINYGLYHPDEIIGKFIKIKGNEYKVTKATNSIINLSNSKNGRVTCLPKRR
VV"

CDS complement(join(15876. .15900,16018. .17390))
/note="GeneID exon scores (in order of location ranges): -0.56, 167.37 - GSCJ_ID dd_00405"
/codon_start=1
/product="similar to Physarum polycephalum (Slime mold). Glutathione reductase"
/protein_id="AAM09354.2"
/db_xref="GI:28828055"
/translation="MSSTNHFTYLVLGAGSGGIASARRAAKHLNAKGNGDRIGIVEVT
RPGGTCVNVGCVPKKVMWNTSFIKEMINAAPSYGFDFGGQQVKFNWPTIKKARDEYIK
RLNGIYDSNLAKDNIVRINGYGRFSGPKEIQVNGANGEKYTADHILIAAGGRPTVPDV
PGKELGITSDGFFELEDLPKSTLVVGAGYIAVELAGVLHSLGSETTMVIRQKQFLRTF
DEMLHTTLLKQMTDDGVKFVTEASIKSLERDVDGKRIIATTNAGVKLPPVECVIWAIG
RVPNTDDLGIDKAGIQLTEQSGFIKVDEFQNTNVPGVHAVGDICGNFLLTPVAIAAGR
RLSERLFNGKSDLKFEYENVATVVFSHPPIGTVGLTEQEAITKYGTENIKCYNTSFIN
MFYSVQVHKVRTSMKLVCLGKEEKVIGLHIIGDGCDEIIQGFAVAVKMGCTKWDLDNT
CAIHPTSAEELVTMV"

CDS complement(join(18106. .18501,18638. .18994))
/note="GeneID exon scores (in order of location ranges): 56.32, 49.02 - GSCJ_ID dd_00404"
/codon_start=1

Alignment Scores:

Pred. No.:	4.48	Length:	254733
Score:	84.00	Matches:	12
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-15 (1-12) x AC117075 (1-254733)

Qy 1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12

Db 129205 CACAACCACAACCACAACCACAACCACAACCACAAC 129240

RESULT 108
AC107178
LOCUS AC107178 255032 bp DNA linear HTG 08-OCT-2002
DEFINITION Rattus norvegicus clone CH230-162F7, *** SEQUENCING IN PROGRESS ***, 2 unordered pieces.
ACCESSION AC107178
VERSION AC107178.4 GI:22856980
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 255032)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,

Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 255032)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 255032)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (08-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Sep 14, 2002 this sequence version replaced gi:21737498.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GPGF
Center clone name: CH230-162F7
---------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 207373 bases at least Q40
Consensus quality: 211449 bases at least Q30
Consensus quality: 212846 bases at least Q20
Estimated insert size: 231056; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 250026: contig of 250026 bp in length
* 250027 250126: gap of unknown length
* 250127 255032: contig of 4906 bp in length.
FEATURES Location/Qualifiers
source 1. .255032
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-162F7"
misc_feature 1. .1126
/note="wgs_contig"
gap 250027. .250126
/estimated_length=unknown
ORIGIN

Alignment Scores:

Pred. No.:	4.48	Length:	255032
Score:	84.00	Matches:	12
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	14	Gaps:	0

US-09-858-332G-15 (1-12) x AC107178 (1-255032)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     249554 CATAATCATAATCATAATCATAATCATAATCATAAT 249589
```

RESULT 109
AC112305
LOCUS AC112305 255050 bp DNA linear HTG 15-NOV-2002
DEFINITION Rattus norvegicus clone CH230-238F22, WORKING DRAFT SEQUENCE.
ACCESSION AC112305
VERSION AC112305.4 GI:25007387
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 255050)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,

```
          Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
          Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
          Weinstock,G. and Gibbs,R.A.
  TITLE   Direct Submission
  JOURNAL Unpublished
REFERENCE 2  (bases 1 to 255050)
  AUTHORS Worley,K.C.
  TITLE   Direct Submission
  JOURNAL Submitted (21-FEB-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3  (bases 1 to 255050)
  AUTHORS Rat Genome Sequencing Consortium.
  TITLE   Direct Submission
  JOURNAL Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Nov 15, 2002 this sequence version replaced gi:23603917.
          The sequence in this assembly is a combination of BAC based reads
          and whole genome shotgun sequencing reads assembled using Atlas
          (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
          in the feature table below represents a scaffold in the Atlas
          assembly (a 'contig-scaffold'). Within each contig-scaffold,
          individual sequence contigs are ordered and oriented, and separated
          by sized gaps filled with Ns to the estimated size. The sequence
          may extend beyond the ends of the clone and there may be sequence
          contigs within a contig-scaffold that consist entirely of whole
          genome shotgun sequence reads. Both end sequences and whole genome
          shotgun sequence only contigs will be indicated in the feature
          table.
          -------------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
          -------------- Project Information
              Center project name: GMYJ
              Center clone name: CH230-238F22
          -------------- Summary Statistics
              Assembly program: Phrap; version 0.990329
              Consensus quality: 231810 bases at least Q40
              Consensus quality: 234330 bases at least Q30
              Consensus quality: 235896 bases at least Q20
              Estimated insert size: 234491; sum-of-contigs estimation
              Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
          --------------
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 1 contigs. Gaps between the contigs
          * are represented as runs of N. The order of the pieces
          * is believed to be correct as given, however the sizes
          * of the gaps between them are based on estimates that have
          * provided by the submittor.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          *       1   255050: contig of 255050 bp in length.
FEATURES             Location/Qualifiers
     source          1. .255050
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-238F22"
     misc_feature    complement(1253. .2114)
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:BZ102269"
ORIGIN

Alignment Scores:
Pred. No.:              4.48          Length:        255050
```

```
Score:                  84.00         Matches:       12
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     14            Gaps:          0

US-09-858-332G-15 (1-12) x AC112305 (1-255050)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     234543 CATAATCATAATCATAATCATAATCATAATCACAAC 234578

RESULT 110
AC151980/c
LOCUS       AC151980              256033 bp   DNA     linear   HTG 19-MAR-2005
DEFINITION  Mus musculus chromosome 8 clone RP24-83M3, WORKING DRAFT SEQUENCE,
            8 unordered pieces.
ACCESSION   AC151980
VERSION     AC151980.2  GI:61656427
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 256033)
  AUTHORS   Wilson,R.K.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 256033)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-2004) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 256033)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAR-2005) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Mar 19, 2005 this sequence version replaced gi:54292022.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu
            Contact: submissions@watson.wustl.edu
            ---------- Project Information ----------
            Center project name: M_BB0083M03
            ----------
            ---------- Summary Statistics  ----------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 249788 bases at least Q40
            Consensus quality: 250953 bases at least Q30
            Consensus quality: 251744 bases at least Q20
            ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1203: contig of 1203 bp in length
            *    1204     1303: gap of unknown length
            *    1304     2565: contig of 1262 bp in length
            *    2566     2665: gap of unknown length
            *    2666     5048: contig of 2383 bp in length
            *    5049     5148: gap of unknown length
```

```
            *     5149    16636: contig of 11488 bp in length
            *    16637    16736: gap of unknown length
            *    16737    20850: contig of 4114 bp in length
            *    20851    20950: gap of unknown length
            *    20951    79081: contig of 58131 bp in length
            *    79082    79181: gap of unknown length
            *    79182   153805: contig of 74624 bp in length
            *   153806   153905: gap of unknown length
            *   153906   256033: contig of 102128 bp in length.
FEATURES             Location/Qualifiers
     source          1. .256033
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="8"
                     /clone="RP24-83M3"
     misc_feature    1. .1203
                     /note="assembly_name:Contig49"
     gap             1204. .1303
                     /estimated_length=unknown
     misc_feature    1304. .2565
                     /note="assembly_name:Contig50"
     gap             2566. .2665
                     /estimated_length=unknown
     misc_feature    2666. .5048
                     /note="assembly_name:Contig57"
     gap             5049. .5148
                     /estimated_length=unknown
     misc_feature    5149. .16636
                     /note="assembly_name:Contig58"
     gap             16637. .16736
                     /estimated_length=unknown
     misc_feature    16737. .20850
                     /note="assembly_name:Contig59"
     gap             20851. .20950
                     /estimated_length=unknown
     misc_feature    20951. .79081
                     /note="assembly_name:Contig60"
     gap             79082. .79181
                     /estimated_length=unknown
     misc_feature    79182. .153805
                     /note="assembly_name:Contig61"
     gap             153806. .153905
                     /estimated_length=unknown
     misc_feature    153906. .256033
                     /note="assembly_name:Contig62"
ORIGIN

Alignment Scores:
Pred. No.:              4.5             Length:         256033
Score:                  84.00           Matches:        12
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-15 (1-12) x AC151980 (1-256033)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      12898 CATAACCATAACCATAACCATAACCATAACCATAAC 12863

RESULT 111
AC114024/c
LOCUS       AC114024               256717 bp    DNA     linear   HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-30O3, *** SEQUENCING IN PROGRESS ***,
            4 unordered pieces.
ACCESSION   AC114024
VERSION     AC114024.4  GI:30581646
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 256717)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 256717)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 256717)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:23268949.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
```

```
          by sized gaps filled with Ns to the estimated size. The sequence
          may extend beyond the ends of the clone and there may be sequence
          contigs within a contig-scaffold that consist entirely of whole
          genome shotgun sequence reads. Both end sequences and whole genome
          shotgun sequence only contigs will be indicated in the feature
          table.
          -------------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
          -------------- Project Information
              Center project name: GKLD
              Center clone name: CH230-3003
          -------------- Summary Statistics
              Assembly program: Atlas 3.0;
              Consensus quality: 222948 bases at least Q40
              Consensus quality: 227438 bases at least Q30
              Consensus quality: 230894 bases at least Q20
              Estimated insert size: 241816; sum-of-contigs estimation
              Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
          --------------
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 4 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1    10694: contig of 10694 bp in length
          *    10695    10794: gap of unknown length
          *    10795   254141: contig of 243347 bp in length
          *   254142   254241: gap of unknown length
          *   254242   255436: contig of 1195 bp in length
          *   255437   255536: gap of unknown length
          *   255537   256717: contig of 1181 bp in length.
FEATURES             Location/Qualifiers
     source          1. .256717
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-3003"
     misc_feature    1. .1268
                     /note="wgs_contig"
     gap             10695. .10794
                     /estimated_length=unknown
     misc_feature    156131. .157293
                     /note="wgs_contig"
     misc_feature    167032. .168923
                     /note="wgs_contig"
     misc_feature    169633. .172165
                     /note="wgs_contig"
     misc_feature    247544. .249982
                     /note="wgs_contig"
     misc_feature    251680. .254141
                     /note="wgs_contig"
     gap             254142. .254241
                     /estimated_length=unknown
     gap             255437. .255536
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              4.51          Length:         256717
Score:                  84.00         Matches:        12
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     14            Gaps:           0

US-09-858-332G-15 (1-12) x AC114024 (1-256717)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      70661 CACAACCATAACCACAACCACAACCACAACCACAAC 70626

RESULT 112
AC136422
LOCUS       AC136422              258253 bp    DNA     linear   HTG 09-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-4F23, *** SEQUENCING IN PROGRESS ***,
            4 unordered pieces.
ACCESSION   AC136422
VERSION     AC136422.3  GI:30466941
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 258253)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 258253)
  AUTHORS   Rat Genome Sequencing Consortium.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (01-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 258253)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 9, 2003 this sequence version replaced gi:25188897.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            ---------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            ---------- Project Information
                Center project name: GAQD
                Center clone name: CH230-4F23
            ---------- Summary Statistics
                Assembly program: Atlas;
                Consensus quality: 217292 bases at least Q40
                Consensus quality: 219787 bases at least Q30
                Consensus quality: 221693 bases at least Q20
                Estimated insert size: 227650; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            ----------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   250922: contig of 250922 bp in length
            *  250923   251022: gap of unknown length
            *  251023   255207: contig of 4185 bp in length
            *  255208   255307: gap of unknown length
            *  255308   256355: contig of 1048 bp in length
            *  256356   256455: gap of unknown length
            *  256456   258253: contig of 1798 bp in length.
FEATURES             Location/Qualifiers
     source          1. .258253
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-4F23"
     misc_feature    1. .1098
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    1916. .3925
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    3976. .5584
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    8979. .9836
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH305446"
     misc_feature    87450. .88730
                     /note="wgs_contig"
     misc_feature    242562. .250462
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH305448"
     gap             250923. .251022
                     /estimated_length=unknown
     gap             255208. .255307
                     /estimated_length=unknown
     gap             256356. .256455
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:                4.54          Length:          258253
Score:                    84.00         Matches:         12
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       14            Gaps:            0

US-09-858-332G-15 (1-12) x AC136422 (1-258253)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     202096 CACAACCACAACCACAACCACAACCACAACCACAAC 202131

RESULT 113
AC119447
LOCUS       AC119447              260849 bp    DNA     linear   HTG 22-SEP-2002
DEFINITION  Rattus norvegicus clone CH230-47O2, *** SEQUENCING IN PROGRESS ***,
            5 unordered pieces.
ACCESSION   AC119447
VERSION     AC119447.4  GI:23267362
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 260849)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
```

Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 260849)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (27-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 260849)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (22-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Sep 22, 2002 this sequence version replaced gi:21746538.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequening reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
-------------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-------------- Project Information
Center project name: GUPC
Center clone name: CH230-47O2
-------------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 228568 bases at least Q40
Consensus quality: 232238 bases at least Q30
Consensus quality: 234535 bases at least Q20
Estimated insert size: 255181; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 20701: contig of 20701 bp in length
* 20702 20801: gap of unknown length
* 20802 41041: contig of 20240 bp in length
* 41042 41141: gap of unknown length
* 41142 156928: contig of 115787 bp in length
* 156929 157028: gap of unknown length
* 157029 235838: contig of 78810 bp in length
* 235839 235938: gap of unknown length
* 235939 260849: contig of 24911 bp in length.

FEATURES Location/Qualifiers
source 1. .260849
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-47O2"
misc_feature 1. .1016
/note="wgs_end_extension
clone_end:T7"
misc_feature 695. .1112
/note="clone_boundary
clone_end:T7
site:EcoRI
end_sequence:BH361669"
misc_feature complement(8162. .8721)
/note="clone_boundary
clone_end:Sp6
site:EcoRI
end_sequence:BH361671"
gap 20702. .20801
/estimated_length=unknown
gap 41042. .41141
/estimated_length=unknown
misc_feature 96543. .98281
/note="wgs_end_extension
clone_end:Sp6"
gap 156929. .157028
/estimated_length=unknown
gap 235839. .235938
/estimated_length=unknown
ORIGIN

Alignment Scores:

Pred. No.:	4.58	Length:	260849
Score:	84.00	Matches:	12
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	14	Gaps:	0

US-09-858-332G-15 (1-12) x AC119447 (1-260849)

Qy 1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
||||||||||||||||||||||||||||||||||||
Db 38590 CATAATCATAATCATAATCATAATCATAATCATAAT 38625

RESULT 114
AC102976
LOCUS AC102976 262405 bp DNA linear HTG 22-SEP-2002
DEFINITION Rattus norvegicus clone CH230-39J8, *** SEQUENCING IN PROGRESS ***, 5 unordered pieces.
ACCESSION AC102976
VERSION AC102976.5 GI:23096569
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 262405)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,

Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 262405)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (24-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 262405)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (22-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Sep 18, 2002 this sequence version replaced gi:22758689.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequening reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GGTM
Center clone name: CH230-39J8

---------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 192222 bases at least Q40
Consensus quality: 197138 bases at least Q30
Consensus quality: 200012 bases at least Q20
Estimated insert size: 220655; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 51246: contig of 51246 bp in length
* 51247 51346: gap of unknown length
* 51347 254883: contig of 203537 bp in length
* 254884 254983: gap of unknown length
* 254984 256476: contig of 1493 bp in length
* 256477 256576: gap of unknown length
* 256577 259335: contig of 2759 bp in length
* 259336 259435: gap of unknown length
* 259436 262405: contig of 2970 bp in length.

FEATURES Location/Qualifiers
source 1. .262405
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-39J8"
misc_feature 14915. .80925
/note="clone_boundary
clone_end:Sp6
site:EcoRI
end_sequence:BH285361"
gap 51247. .51346
/estimated_length=unknown
misc_feature 137042. .138655
/note="wgs_contig"
gap 254884. .254983
/estimated_length=unknown
gap 256477. .256576
/estimated_length=unknown
gap 259336. .259435
/estimated_length=unknown

ORIGIN

Alignment Scores:

Pred. No.:	4.61	Length:	262405
Score:	84.00	Matches:	12
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	14	Gaps:	0

US-09-858-332G-15 (1-12) x AC102976 (1-262405)

Qy 1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
||||||||||||||||||||||||||||||||||||
Db 173702 CACAACCACAACCACAACCACAACCACAACCACAAC 173737

RESULT 115
AC121457
LOCUS AC121457 264615 bp DNA linear HTG 12-OCT-2002
DEFINITION Rattus norvegicus clone CH230-146J5, *** SEQUENCING IN PROGRESS ***, 2 unordered pieces.
ACCESSION AC121457
VERSION AC121457.3 GI:23907899
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)

```
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 264615)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 264615)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 264615)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Oct 12, 2002 this sequence version replaced gi:21908060.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GYKE
                Center clone name: CH230-146J5
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 207625 bases at least Q40
                Consensus quality: 211034 bases at least Q30
                Consensus quality: 213357 bases at least Q20
                Estimated insert size: 215749; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1   263260: contig of 263260 bp in length
            *   263261   263360: gap of unknown length
            *   263361   264615: contig of 1255 bp in length.
FEATURES             Location/Qualifiers
     source          1. .264615
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-146J5"
     misc_feature    1. .1514
                     /note="wgs_contig"
     misc_feature    1565. .3140
                     /note="wgs_contig"
     misc_feature    4867. .6350
                     /note="wgs_contig"
     misc_feature    complement(262294. .263119)
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:RWBAG51TJB"
     gap             263261. .263360
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              4.64          Length:         264615
Score:                  84.00         Matches:        12
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     14            Gaps:           0

US-09-858-332G-15 (1-12) x AC121457 (1-264615)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     161833 CACAACCACAACCACAACCACAACCACAACCACAAC 161868

RESULT 116
```

```
AC110912/c
LOCUS       AC110912              268063 bp    DNA     linear   HTG 23-APR-2003
DEFINITION  Mus musculus clone RP24-397B15, WORKING DRAFT SEQUENCE, 68
            unordered pieces.
ACCESSION   AC110912
VERSION     AC110912.2  GI:30018053
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 268063)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-397B15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 268063)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 268063)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-APR-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 17, 2003 this sequence version replaced gi:18693475.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L21276
                Center clone name: 397_B_15
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 68 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    24290: contig of 24290 bp in length
            *    24291    24390: gap of 100 bp
            *    24391    25057: contig of 667 bp in length
            *    25058    25157: gap of 100 bp
            *    25158    25809: contig of 652 bp in length
            *    25810    25909: gap of 100 bp
            *    25910    26753: contig of 844 bp in length
            *    26754    26853: gap of 100 bp
            *    26854    27553: contig of 700 bp in length
            *    27554    27653: gap of 100 bp
            *    27654    29159: contig of 1506 bp in length
            *    29160    29259: gap of 100 bp
            *    29260    30197: contig of 938 bp in length
            *    30198    30297: gap of 100 bp
            *    30298    31460: contig of 1163 bp in length
            *    31461    31560: gap of 100 bp
            *    31561    32590: contig of 1030 bp in length
            *    32591    32690: gap of 100 bp
            *    32691    33703: contig of 1013 bp in length
            *    33704    33803: gap of 100 bp
            *    33804    35074: contig of 1271 bp in length
            *    35075    35174: gap of 100 bp
            *    35175    36345: contig of 1171 bp in length
            *    36346    36445: gap of 100 bp
            *    36446    37894: contig of 1449 bp in length
            *    37895    37994: gap of 100 bp
            *    37995    39318: contig of 1324 bp in length
            *    39319    39418: gap of 100 bp
            *    39419    40661: contig of 1243 bp in length
            *    40662    40761: gap of 100 bp
            *    40762    42129: contig of 1368 bp in length
            *    42130    42229: gap of 100 bp
            *    42230    43797: contig of 1568 bp in length
            *    43798    43897: gap of 100 bp
            *    43898    44862: contig of 965 bp in length
            *    44863    44962: gap of 100 bp
            *    44963    46854: contig of 1892 bp in length
            *    46855    46954: gap of 100 bp
            *    46955    47927: contig of 973 bp in length
            *    47928    48027: gap of 100 bp
            *    48028    49244: contig of 1217 bp in length
            *    49245    49344: gap of 100 bp
            *    49345    50632: contig of 1288 bp in length
            *    50633    50732: gap of 100 bp
            *    50733    52049: contig of 1317 bp in length
            *    52050    52149: gap of 100 bp
            *    52150    53237: contig of 1088 bp in length
            *    53238    53337: gap of 100 bp
            *    53338    55277: contig of 1940 bp in length
            *    55278    55377: gap of 100 bp
            *    55378    56685: contig of 1308 bp in length
            *    56686    56785: gap of 100 bp
            *    56786    58642: contig of 1857 bp in length
            *    58643    58742: gap of 100 bp
            *    58743    60107: contig of 1365 bp in length
            *    60108    60207: gap of 100 bp
            *    60208    61814: contig of 1607 bp in length
            *    61815    61914: gap of 100 bp
            *    61915    63033: contig of 1119 bp in length
```

```
*     63034     63133: gap of 100 bp
*     63134     64082: contig of 949 bp in length
*     64083     64182: gap of 100 bp
*     64183     66519: contig of 2337 bp in length
*     66520     66619: gap of 100 bp
*     66620     68694: contig of 2075 bp in length
*     68695     68794: gap of 100 bp
*     68795     70253: contig of 1459 bp in length
*     70254     70353: gap of 100 bp
*     70354     72323: contig of 1970 bp in length
*     72324     72423: gap of 100 bp
*     72424     74360: contig of 1937 bp in length
*     74361     74460: gap of 100 bp
*     74461     76252: contig of 1792 bp in length
*     76253     76352: gap of 100 bp
*     76353     78129: contig of 1777 bp in length
*     78130     78229: gap of 100 bp
*     78230     80637: contig of 2408 bp in length
*     80638     80737: gap of 100 bp
*     80738     82870: contig of 2133 bp in length
*     82871     82970: gap of 100 bp
*     82971     85712: contig of 2742 bp in length
*     85713     85812: gap of 100 bp
*     85813     87890: contig of 2078 bp in length
*     87891     87990: gap of 100 bp
*     87991     90163: contig of 2173 bp in length
*     90164     90263: gap of 100 bp
*     90264     92559: contig of 2296 bp in length
*     92560     92659: gap of 100 bp
*     92660     95748: contig of 3089 bp in length
*     95749     95848: gap of 100 bp
*     95849     98374: contig of 2526 bp in length
*     98375     98474: gap of 100 bp
*     98475    101777: contig of 3303 bp in length
*    101778    101877: gap of 100 bp
*    101878    104813: contig of 2936 bp in length
*    104814    104913: gap of 100 bp
*    104914    107515: contig of 2602 bp in length
*    107516    107615: gap of 100 bp
*    107616    110328: contig of 2713 bp in length
*    110329    110428: gap of 100 bp
*    110429    114176: contig of 3748 bp in length
*    114177    114276: gap of 100 bp
*    114277    118279: contig of 4003 bp in length
*    118280    118379: gap of 100 bp
*    118380    122029: contig of 3650 bp in length
*    122030    122129: gap of 100 bp
*    122130    126278: contig of 4149 bp in length
*    126279    126378: gap of 100 bp
*    126379    130335: contig of 3957 bp in length
*    130336    130435: gap of 100 bp
*    130436    134925: contig of 4490 bp in length
*    134926    135025: gap of 100 bp
*    135026    138675: contig of 3650 bp in length
*    138676    138775: gap of 100 bp

Alignment Scores:
Pred. No.:                4.7          Length:          268063
Score:                    84.00        Matches:         12
Percent Similarity:       100.00%      Conservative:    0
Best Local Similarity:    100.00%      Mismatches:      0
Query Match:              100.00%      Indels:          0
DB:                       14           Gaps:            0

US-09-858-332G-15 (1-12) x AC110912 (1-268063)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     250744 CATAATCATAATCATAATCATAACCATAATCATAAT 250709

RESULT 117
AC091323
LOCUS       AC091323               268644 bp    DNA     linear   HTG 03-MAR-2004
DEFINITION  Mus musculus chromosome 17 clone RP23-333E5 map 17, *** SEQUENCING
            IN PROGRESS ***, 8 unordered pieces.
ACCESSION   AC091323
VERSION     AC091323.4  GI:39752830
KEYWORDS    HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 268644)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 17, clone RP23-333E5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 268644)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
            Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K.,
            Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
            Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-APR-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 268644)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-MAR-2004) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Dec 12, 2003 this sequence version replaced gi:24182393.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
```

```
            Center project name: L13287
            Center clone name: 333_E_5
          ----------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 8 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *       1    82781: contig of 82781 bp in length
          *   82782    82881: gap of 100 bp
          *   82882   126562: contig of 43681 bp in length
          *  126563   126662: gap of 100 bp
          *  126663   134445: contig of 7783 bp in length
          *  134446   134545: gap of 100 bp
          *  134546   142101: contig of 7556 bp in length
          *  142102   142201: gap of 100 bp
          *  142202   203346: contig of 61145 bp in length
          *  203347   203446: gap of 100 bp
          *  203447   235687: contig of 32241 bp in length
          *  235688   235787: gap of 100 bp
          *  235788   251563: contig of 15776 bp in length
          *  251564   251663: gap of 100 bp
          *  251664   268644: contig of 16981 bp in length.
FEATURES             Location/Qualifiers
     source          1. .268644
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="17"
                     /map="17"
                     /clone="RP23-333E5"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     gap             82782. .82881
                     /estimated_length=100
     gap             126563. .126662
                     /estimated_length=100
     gap             134446. .134545
                     /estimated_length=100
     gap             142102. .142201
                     /estimated_length=100
     gap             203347. .203446
                     /estimated_length=100
     gap             235688. .235787
                     /estimated_length=100
     gap             251564. .251663
                     /estimated_length=100
ORIGIN

Alignment Scores:
Pred. No.:              4.71          Length:         268644
Score:                 84.00          Matches:        12
Percent Similarity:   100.00%         Conservative:   0
Best Local Similarity: 100.00%        Mismatches:     0
Query Match:           100.00%        Indels:         0
DB:                    14             Gaps:           0

US-09-858-332G-15 (1-12) x AC091323 (1-268644)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     263920 CACAACCACAACCACAACCACAACCACAACCACAAC 263955

RESULT 118
AC133056
LOCUS       AC133056               269740 bp    DNA     linear   HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-25D4, WORKING DRAFT SEQUENCE, 6
            unordered pieces.
ACCESSION   AC133056
VERSION     AC133056.4  GI:30581349
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 269740)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 269740)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 269740)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:23112649.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
```

in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GKJA
Center clone name: CH230-25D4
---------- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 237706 bases at least Q40
Consensus quality: 240031 bases at least Q30
Consensus quality: 241569 bases at least Q20
Estimated insert size: 243784; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
---
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 10150: contig of 10150 bp in length
* 10151 10250: gap of unknown length
* 10251 13830: contig of 3580 bp in length
* 13831 13930: gap of unknown length
* 13931 266092: contig of 252162 bp in length
* 266093 266192: gap of unknown length
* 266193 267411: contig of 1219 bp in length
* 267412 267511: gap of unknown length
* 267512 268555: contig of 1044 bp in length
* 268556 268655: gap of unknown length
* 268656 269740: contig of 1085 bp in length.

FEATURES Location/Qualifiers
source 1. .269740
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-25D4"
misc_feature 1. .1131
/note="wgs_contig"
misc_feature 6403. .8291
/note="wgs_contig"
gap 10151. .10250
/estimated_length=unknown
gap 13831. .13930
/estimated_length=unknown
misc_feature 13931. .15418
/note="wgs_contig"
misc_feature 16612. .20299
/note="wgs_contig"
misc_feature 263153. .264945
/note="wgs_contig"
misc_feature 264996. .266092
/note="wgs_contig"
gap 266093. .266192
/estimated_length=unknown
gap 267412. .267511
/estimated_length=unknown
gap 268556. .268655
/estimated_length=unknown

ORIGIN

Alignment Scores:

Pred. No.:	4.73	Length:	269740
Score:	84.00	Matches:	12
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	14	Gaps:	0

US-09-858-332G-15 (1-12) x AC133056 (1-269740)

Qy 1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
||||||||||||||||||||||||||||||||||||
Db 253433 CACAACCACAACCACAACCACAACCACAACCACAAC 253468

RESULT 119
AC106147
LOCUS AC106147 273729 bp DNA linear HTG 13-MAY-2003
DEFINITION Rattus norvegicus clone CH230-185A16, WORKING DRAFT SEQUENCE, 4 unordered pieces.
ACCESSION AC106147
VERSION AC106147.5 GI:30579549
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 273729)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,

```
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 273729)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 273729)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:24819222.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GJHA
                Center clone name: CH230-185A16
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 226011 bases at least Q40
                Consensus quality: 228238 bases at least Q30
                Consensus quality: 229717 bases at least Q20
                Estimated insert size: 234596; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   269813: contig of 269813 bp in length
            *  269814   269913: gap of unknown length
            *  269914   270928: contig of 1015 bp in length
            *  270929   271028: gap of unknown length
            *  271029   272482: contig of 1454 bp in length
            *  272483   272582: gap of unknown length
            *  272583   273729: contig of 1147 bp in length.
FEATURES                Location/Qualifiers
     source             1. .273729
                        /organism="Rattus norvegicus"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:10116"
                        /clone="CH230-185A16"
     misc_feature       1. .2958
                        /note="wgs_end_extension
                        clone_end:Sp6"
     misc_feature       3009. .4394
                        /note="wgs_end_extension
                        clone_end:Sp6"
     misc_feature       4997. .5807
                        /note="clone_boundary
                        clone_end:Sp6
                        site:EcoRI
                        end_sequence:BH261673"
     misc_feature       223544. .224746
                        /note="wgs_contig"
     misc_feature       225798. .226894
                        /note="wgs_contig"
     misc_feature       complement(227890. .228685)
                        /note="clone_boundary
                        clone_end:T7
                        site:EcoRI
                        end_sequence:BH261671"
     gap                269814. .269913
                        /estimated_length=unknown
     gap                270929. .271028
                        /estimated_length=unknown
     gap                272483. .272582
                        /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              4.8             Length:         273729
Score:                  84.00           Matches:        12
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-15 (1-12) x AC106147 (1-273729)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     185125 CACAACCACAACCACAACCACAACCACAACCACAAC 185160

RESULT 120
AC115186/c
LOCUS       AC115186               274368 bp    DNA     linear   HTG 08-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-152P11, *** SEQUENCING IN PROGRESS
            ***, 3 unordered pieces.
ACCESSION   AC115186
VERSION     AC115186.3  GI:22856601
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 274368)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
```

```
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 274368)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 274368)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 14, 2002 this sequence version replaced gi:21745587.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GSUI
                Center clone name: CH230-152P11
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 200619 bases at least Q40
                Consensus quality: 203437 bases at least Q30
                Consensus quality: 205038 bases at least Q20
                Estimated insert size: 218032; sum-of-contigs estimation
                Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   262593: contig of 262593 bp in length
            *  262594   262693: gap of unknown length
            *  262694   264394: contig of 1701 bp in length
            *  264395   264494: gap of unknown length
            *  264495   274368: contig of 9874 bp in length.
FEATURES             Location/Qualifiers
     source          1. .274368
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-152P11"
     misc_feature    1. .1036
                     /note="wgs_contig"
     misc_feature    1428. .3679
                     /note="wgs_contig"
     misc_feature    261472. .262593
                     /note="wgs_contig"
     gap             262594. .262693
                     /estimated_length=unknown
     gap             264395. .264494
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              4.81          Length:         274368
Score:                  84.00         Matches:        12
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     14            Gaps:           0

US-09-858-332G-15 (1-12) x AC115186 (1-274368)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      98704 CACAACCACAACCACAACCACAACCACAACCACAAC 98669

RESULT 121
AC126663
LOCUS       AC126663              286094 bp    DNA     linear   HTG 10-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-4C9, *** SEQUENCING IN PROGRESS ***,
            4 unordered pieces.
ACCESSION   AC126663
VERSION     AC126663.5  GI:22771616
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 286094)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
```

Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 286094)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (08-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 286094)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (10-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Sep 10, 2002 this sequence version replaced gi:22759261.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GBIN
Center clone name: CH230-4C9
---------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 216918 bases at least Q40
Consensus quality: 220868 bases at least Q30
Consensus quality: 223082 bases at least Q20
Estimated insert size: 227119; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 249362: contig of 249362 bp in length
* 249363 249462: gap of unknown length
* 249463 283417: contig of 33955 bp in length
* 283418 283517: gap of unknown length
* 283518 284542: contig of 1025 bp in length
* 284543 284642: gap of unknown length
* 284643 286094: contig of 1452 bp in length.

FEATURES Location/Qualifiers
source 1. .286094
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-4C9"
gap 249363. .249462
/estimated_length=unknown
gap 283418. .283517
/estimated_length=unknown
gap 284543. .284642
/estimated_length=unknown

ORIGIN

Alignment Scores:

Pred. No.:	5.01	Length:	286094
Score:	84.00	Matches:	12
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	14	Gaps:	0

US-09-858-332G-15 (1-12) x AC126663 (1-286094)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      67273 CATAATCATAATCATAATCATAATCATAATCATAAT 67308
```

RESULT 122
AC103294
LOCUS AC103294 287586 bp DNA linear HTG 13-MAY-2003
DEFINITION Rattus norvegicus clone CH230-142D22, *** SEQUENCING IN PROGRESS ***, 10 unordered pieces.
ACCESSION AC103294
VERSION AC103294.6 GI:30579163
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 287586)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,

```
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 287586)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-NOV-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 287586)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:23269031.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GJTO
                Center clone name: CH230-142D22
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 247075 bases at least Q40
                Consensus quality: 249492 bases at least Q30
                Consensus quality: 251363 bases at least Q20
                Estimated insert size: 256609; sum-of-contigs estimation
                Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    10722: contig of 10722 bp in length
            *    10723    10822: gap of unknown length
            *    10823    21877: contig of 11055 bp in length
            *    21878    21977: gap of unknown length
            *    21978   273878: contig of 251901 bp in length
            *   273879   273978: gap of unknown length
            *   273979   277637: contig of 3659 bp in length
            *   277638   277737: gap of unknown length
            *   277738   278747: contig of 1010 bp in length
            *   278748   278847: gap of unknown length
            *   278848   280061: contig of 1214 bp in length
            *   280062   280161: gap of unknown length
            *   280162   281487: contig of 1326 bp in length
            *   281488   281587: gap of unknown length
            *   281588   282774: contig of 1187 bp in length
            *   282775   282874: gap of unknown length
            *   282875   284477: contig of 1603 bp in length
            *   284478   284577: gap of unknown length
            *   284578   287586: contig of 3009 bp in length.
FEATURES             Location/Qualifiers
     source          1. .287586
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-142D22"
     misc_feature    1. .2072
                     /note="wgs_contig"
     gap             10723. .10822
                     /estimated_length=unknown
     misc_feature    10823. .12542
                     /note="wgs_contig"
     gap             21878. .21977
                     /estimated_length=unknown
     misc_feature    21978. .23465
                     /note="wgs_contig"
     misc_feature    270939. .272731
                     /note="wgs_contig"
     misc_feature    272782. .273878
                     /note="wgs_contig"
     gap             273879. .273978
                     /estimated_length=unknown
     gap             277638. .277737
                     /estimated_length=unknown
     gap             278748. .278847
                     /estimated_length=unknown
     gap             280062. .280161
                     /estimated_length=unknown
     gap             281488. .281587
```

```
                     /estimated_length=unknown
     gap             282775. .282874
                     /estimated_length=unknown
     gap             284478. .284577
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:                5.03          Length:          287586
Score:                    84.00         Matches:         12
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       14            Gaps:            0

US-09-858-332G-15 (1-12) x AC103294 (1-287586)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     261219 CACAACCACAACCACAACCACAACCACAACCACAAC 261254

RESULT 123
AC120291
LOCUS       AC120291               295223 bp    DNA     linear   HTG 20-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-21G15, *** SEQUENCING IN PROGRESS
            ***, 9 unordered pieces.
ACCESSION   AC120291
VERSION     AC120291.4  GI:25137993
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 295223)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 295223)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 295223)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 20, 2002 this sequence version replaced gi:23269633.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GWWM
                Center clone name: CH230-21G15
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 210290 bases at least Q40
                Consensus quality: 214669 bases at least Q30
                Consensus quality: 217523 bases at least Q20
                Estimated insert size: 216345; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    17319: contig of 17319 bp in length
            *    17320    17419: gap of unknown length
            *    17420    28564: contig of 11145 bp in length
            *    28565    28664: gap of unknown length
            *    28665    64138: contig of 35474 bp in length
            *    64139    64238: gap of unknown length
            *    64239   262013: contig of 197775 bp in length
            *   262014   262113: gap of unknown length
            *   262114   276545: contig of 14432 bp in length
```

```
            *   276546   276645: gap of unknown length
            *   276646   278035: contig of 1390 bp in length
            *   278036   278135: gap of unknown length
            *   278136   279240: contig of 1105 bp in length
            *   279241   279340: gap of unknown length
            *   279341   280489: contig of 1149 bp in length
            *   280490   280589: gap of unknown length
            *   280590   295223: contig of 14634 bp in length.
FEATURES             Location/Qualifiers
     source          1. .295223
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-21G15"
     misc_feature    1. .1576
                     /note="wgs_contig"
     gap             17320. .17419
                     /estimated_length=unknown
     gap             28565. .28664
                     /estimated_length=unknown
     gap             64139. .64238
                     /estimated_length=unknown
     gap             262014. .262113
                     /estimated_length=unknown
     misc_feature    262114. .263495
                     /note="wgs_contig"
     gap             276546. .276645
                     /estimated_length=unknown
     gap             278036. .278135
                     /estimated_length=unknown
     gap             279241. .279340
                     /estimated_length=unknown
     gap             280490. .280589
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              5.16          Length:         295223
Score:                  84.00         Matches:        12
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     14            Gaps:           0

US-09-858-332G-15 (1-12) x AC120291 (1-295223)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db      75110 CACAACCACAACCACAACCACAACCACAACCACAAC 75145

RESULT 124
AE017165
LOCUS       AE017165               300143 bp    DNA     linear   BCT 14-AUG-2003
DEFINITION  Prochlorococcus marinus subsp. marinus str. CCMP1375 section 5 of 6
            of the complete genome.
ACCESSION   AE017165 AE017126
VERSION     AE017165.1  GI:33238289
KEYWORDS    .
SOURCE      Prochlorococcus marinus subsp. marinus str. CCMP1375
            (Prochlorococcus marinus SS120)
  ORGANISM  Prochlorococcus marinus subsp. marinus str. CCMP1375
            Bacteria; Cyanobacteria; Prochlorales; Prochlorococcaceae;
            Prochlorococcus.
REFERENCE   1  (bases 1 to 300143)
  AUTHORS   Dufresne,A., Salanoubat,M., Partensky,F., Artiguenave,F.,
            Axmann,I.M., Barbe,V., Duprat,S., Galperin,M.Y., Koonin,E.V., Le
            Gall,F., Makarova,K.S., Ostrowski,M., Oztas,S., Robert,C.,
            Rogozin,I.B., Scanlan,D.J., Tandeau de Marsac,N., Weissenbach,J.,
            Wincker,P., Wolf,Y.I. and Hess,W.R.
  TITLE     Genome sequence of the cyanobacterium Prochlorococcus marinus
            SS120, a near minimal oxyphototrophic genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 100 (17), 10020-10025 (2003)
REFERENCE   2  (bases 1 to 300143)
  AUTHORS   Dufresne,A., Salanoubat,M., Partensky,F., Artiguenave,F.,
            Axmann,I.M., Barbe,V., Duprat,S., Galperin,M.Y., Koonin,E.V., Le
            Gall,F., Makarova,K.S., Ostrowski,M., Oztas,S., Robert,C.,
            Rogozin,I.B., Scanlan,D.J., Tandeau de Marsac,N., Weissenbach,J.,
            Wincker,P., Wolf,Y.I. and Hess,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAY-2003) UMR 7127 CNRS et Universite Pierre et Marie
            Curie, Station Biologique de Roscoff, Place Georges Tessier BP74,
            Roscoff 29682, France
FEATURES             Location/Qualifiers
     source          1. .300143
                     /organism="Prochlorococcus marinus subsp. marinus str.
                     CCMP1375"
                     /mol_type="genomic DNA"
                     /strain="CCMP1375; SS120"
                     /sub_species="marinus"
                     /db_xref="taxon:167539"
     tRNA            420. .493
                     /product="tRNA-Arg"
                     /note="codons recognized: AGR"
     gene            complement(509. .1924)
                     /gene="ugd"
                     /locus_tag="Pro1312"
     CDS             complement(509. .1924)
                     /gene="ugd"
                     /locus_tag="Pro1312"
                     /codon_start=1
                     /transl_table=11
                     /product="UDP-glucose 6-dehydrogenase"
                     /protein_id="AAQ00356.1"
                     /db_xref="GI:33238290"
                     /translation="MLEKINNICCIGAGYVGGPTMAVIADKCPEIQVNVVDINQARIN
                     NWNDSDFSNLPIYEPGLAEIVKNCRGRNLHFSTVIEENIALADMIFISVNTPVKKKGI
                     GAGQASDLKWIEASARQISKYAVGKTIVVEKSTLPVKTAQTIKNILEAHVLNTSNDKK
                     FSILSNPEFLAEGTAINDLENPDRVLIGGEDQDAINLLIDVYLNWVDKEKILTTDLWS
                     SELSKLTANAFLAQRISSINSISALCESTGADVNDVALAIGMDKRIGLEFLRPGPGFG
                     GSCFPKKDILNLVYICNHYGLYQAAKYWQTVINLNDWQQKRISKIIVEKLFGTISGKKI
                     AVLGFAFKANTNDTRESPAINICRDLLEEGSNLHIYDPRVSQDQIKMDLGHSQIIDSQ
                     NLLFEGKWEFSKSVELAANGADAILVLTDWEEFKTLDWEKLSKIMRSPSWLFDTRSIS
                     NAIEAKSFGINIWRLGDGALN"
     gene            complement(1928. .2851)
                     /gene="wcaG"
                     /locus_tag="Pro1313"
     CDS             complement(1928. .2851)
                     /gene="wcaG"
                     /locus_tag="Pro1313"
                     /codon_start=1
                     /transl_table=11
                     /product="NAD dependent epimerase/dehydratase"
                     /protein_id="AAQ00357.1"
                     /db_xref="GI:33238291"
                     /translation="MKNLVTGGAGFIGSHVVDRLMNCGEKVICLDNLCTGSLENIKTW
                     IDNPNFQLINHDVINPIELNVDRIWHLACPASPLHYQENPIKTAKTSFLGTYNMLGMA
                     RRTKARLLFASTSEVYGDPEIHPQPETYNGSVNPTQIRSCYTEGKRIAESLCFDYLRE
                     HKLEIRVARIFNTYGPRMLPNDGRVISNFISQAIAKRPHTIYGDGLQTRSFCYVDDLV
                     DALIRLMLSNCSGPINLGNPQECTILELSRIISKKINATYDFITFSLPKDDPMRRKPD
                     INLAKRELDWEPLINLDQGLNLTIDYFKGEL"
     gene            complement(2973. .4016)
                     /gene="galE"
                     /locus_tag="Pro1314"
     CDS             complement(2973. .4016)
                     /gene="galE"
                     /locus_tag="Pro1314"
                     /codon_start=1
                     /transl_table=11
                     /product="UDP-glucose 4-epimerase"
                     /protein_id="AAQ00358.1"
                     /db_xref="GI:33238292"
                     /translation="MNRILVTGGAGFIGSHTCITLLESGYDLLIIDSFINSSEVSLKR
                     VKEMSDNKNINYLKGDIRNLEFLDSIFSNAIQEGKPIDAVIHFAGLKSVSESTKIPLT
                     YWDVNVVGSIALFSIMRKYKCYTIVFSSSATIYGNTDKVPIKEDSLISPINPYGESKA
                     TVEKILSDLSLSAPFDWRIACLRYFNPVGAHPSGRIGEDPLGIPNNLFPYITNVAGGQ
                     IKQVEVFGNDWPTQDGTGVRDYVHVLDLAEAHKSALECLFAEPAQLLILNLGNGFGLS
```

```
                VLEIINTFSRVNNCEVPYVFAARRPGDIAISYADIALSKARLNWYPKRSIEDMCRDTW
                RWKLNNPIGYRSK"
     gene       complement(4208. .5233)
                /gene="wcaG"
                /locus_tag="Pro1315"
     CDS        complement(4208. .5233)
                /gene="wcaG"
                /locus_tag="Pro1315"
                /codon_start=1
                /transl_table=11
                /product="NAD dependent epimerase/dehydratase"
                /protein_id="AAQ00359.1"
                /db_xref="GI:33238293"
                /translation="MRPVLITGAAGFIGAALVKKLLSEGERVIGIDNINEYYDTSLKY
                ARLQYIESSSQEYLKNWVFHKLGIENIDEIKEIFTKESPRIVVNLAAQAGVRYSLDNP
                HAYVQSNLVGFCNLLEMCRHYEVENLIYASSSSVYGGNTNLPFNETQAVNHPVSFYAA
                TKKSNELMAHTYSHLYDLPATGLRFFTVYGPWGRPDMAPMKFAKAILEGDPIQVFNFG
                RMKRDFTYIDDVVESIFRCCYLQATVNNNFDKSNPDPCTSFAPHRIFNVGNSQPTELD
                IFINLLEKSLSTKAIRDLLPMQPGDVVATAADTQKLKDWTGFSPSTSLEDGIQCFAEW
                YLSFYQS"
     gene       5464. .6678
                /gene="wza"
                /locus_tag="Pro1316"
     CDS        5464. .6678
                /gene="wza"
                /locus_tag="Pro1316"
                /note="involved in polysaccharide export"
                /codon_start=1
                /transl_table=11
                /product="polysaccharide export-related periplasmic
                protein"
                /protein_id="AAQ00360.1"
                /db_xref="GI:33238294"
                /translation="MLLKKLSKNIAIILVFTHSICSSVYSIESKKNEEVINNQPSIDY
                LDKIPKGEYLLGPGDSITIIFNSASKLSTRQTVDGDGTIFTSRLKRIYVEGLSINELT
                KLLERKYSEFLIEPNISILVTTYRPINIYVQGEVQNPGVYLIKGGEINLSLGNVNDPF
                GDDFLNPSDLTEDQRTLIEKFPDEYDERFQKKKTVYDGPVAPKLFDAIKRAGGITLYS
                DLRSVEVVRKQSISNGGGRIRARLDFLSVLENGEGAENIRLRDGDIIRINKSELALNE
                QLKIATRTNLHSKFNKVFISGRVYYPGAKVINKSATLNDLIVLSGGMKPLKGKVTFVR
                FLSDGTLERRSIRYRKNAKDNTYHNPVLRSGDIVNVEDSLIAKTSVGISEITNPFLGL
                YSTYKVIESITN"
     gene       complement(6693. .7382)
                /locus_tag="Pro1317"
     CDS        complement(6693. .7382)
                /locus_tag="Pro1317"
                /codon_start=1
                /transl_table=11
                /product="Protein distantly related to SAM-dependent
                methyltransferases"
                /protein_id="AAQ00361.1"
                /db_xref="GI:33238295"
                /translation="MNSLLSYFRYLYRLFLPIQLRTAFRDFYKTFLYALRYLKLRFVL
                LNKSNVNLILGAALTSQEGWYSTNENWLDISKTEHWQRIFSGRPRVSHAVAEHVFEHL
                SVSEMRKALRLIYQHLCNGGSLRIAVPDGNHPDETYRKHTGINGIGADASDHKQFISY
                EFLKKELEDVGFTCELKEGYTSEGQLIRSNISRNYGLIMRTRDSINYKTEKLGWDFID
                SNTSLIVDALK"
     gene       7808. .8467
                /gene="hisB"
                /locus_tag="Pro1318"
     CDS        7808. .8467
                /gene="hisB"
                /locus_tag="Pro1318"
                /codon_start=1
                /transl_table=11
                /product="Phosphatase, HAD superfamily"
                /protein_id="AAQ00362.1"
                /db_xref="GI:33238296"
                /translation="MNKDYSNFISLFYDNPLYLPESMGLRPALFLDRDGVIIKDCHFV
                SNQDDVILEKGCKDFFKEAKSLNIPIIIITNQSGISRGYFGWDNYIEVTNKMIELIGE
                ENSIIGIYANGQGPESPLHSWRKPSPNMILNSAITLSVDLSKSLMIGDRISDLQSAAR
                AGIKSLIHIKTGHGKDERSDILSQTDGQHRFIDKNNRSDLILIDNLESFPFYKLSSFN
                E"
     gene       8469. .9695
                /locus_tag="Pro1319"
     CDS        8469. .9695
                /locus_tag="Pro1319"
                /codon_start=1
                /transl_table=11
                /product="FAD dependent oxidoreductase"
                /protein_id="AAQ00363.1"
                /db_xref="GI:33238297"
                /translation="MINSETKDIIIIGGGMVGLSLAYQLIERRITKSITILDKEKALG
                LHTSGRNSGVLHAGIYYKPGSLKAKVSVEGARRLCEWIEERNLAINKCGKVVIPTKAN
                LDCQLDLLKERGELNGAKVELWDENQLKAFIPQAVSKSGRALWSPNTAVVKPLEIIKC
                LERELTSKGVTIKKGIKIINVNTKEKLITTFNKENLSYSYVFNCAGLGADRISKLFDV
                GNDYTIIPFKGIYWKLKHESSIKIPSNLYPVPDLSVPFLGVHFTPDTERVPSIYIGPT
                ATPAWGRENYDGIKGLEPSVAISSIDLLSRQYINNKGGFRGYVHQQALQSFQPFFLKA
                AQELIPSIRFKDIEPCSKRGIRSQLFNKRSMKLEDDFLCINSDNSSHILNAVSPAFTA
                SFALADLIIDSSILLN"
     gene       9700. .10428
                /gene="GCD1"
                /locus_tag="Pro1320"
     CDS        9700. .10428
                /gene="GCD1"
                /locus_tag="Pro1320"
                /codon_start=1
                /transl_table=11
                /product="Nucleotidyl transferase family enzyme"
                /protein_id="AAQ00364.1"
                /db_xref="GI:33238298"
                /translation="MKLPIRALLLAAGFGTRLRPLTLNTPKCLVSISNKPLLHIWLDK
                LVNLGCKSTLINTHYLSDQVNSSIREYDNSKINIYTTYEKTLLGTAGTLMVNRDFFRG
                SLGLIIHADNITNDNLEELIDTHVNKSKDSLLTMLTFKTDNPSQCGIVETNEKGVVTA
                FHEKTKNPPGFIANGAIYAFDQSFLDFLDQTHFSGQDISKDLLPKLLGRIQTCYSQTT
                FLDIGSPTTLSRAQELWKKESLNE"
     gene       10431. .11021

Alignment Scores:
Pred. No.:                5.24          Length:          300143
Score:                    84.00         Matches:         12
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       1             Gaps:            0

US-09-858-332G-15 (1-12) x AE017165 (1-300143)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     173113 CATAACCATAACCATAACCATAACCATAACCATAAC 173148

RESULT 125
CNS09S4S
LOCUS       CNS09S4S              322194 bp    DNA     linear   PLN 17-APR-2005
DEFINITION  DNA centromeric region sequence from BAC DP26B06, DP34F04, DP16D11,
            DP09G08, DP35C12 of chromosome 5  of Podospora anserina.
ACCESSION   BX088700
VERSION     BX088700.1  GI:27802984
KEYWORDS    .
SOURCE      Podospora anserina
  ORGANISM  Podospora anserina
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetidae; Sordariales; Lasiosphaeriaceae; Podospora.
REFERENCE   1  (bases 1 to 322194)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JAN-2003) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     - Robert Debuchy, Sebastien Kicka,Philippe Sila, Institut de
            Genetique et Microbiologie, UMR CNRS 8621, batiment 400, Universite
            Paris-Sud, F-91405 Orsay CEDEX, France;
            - Laurence Cattolico, Simone Duprat, Genoscope, Centre National de
            Sequencage, 2 rue Gaston Cremieux, CP5706, F-91057 Evry cedex,
            France; - Christian Barreau, Beatrice Turcq, Institut de Biochimie
            et de Genetique Cellulaire, UPR CNRS 9026, 1 rue Camille Saint
```

```
          Saens, F-33077 Bordeaux, France;
          - Alain Billault, Molecular Engines Laboratories, 20 rue Bouvier,
          F-75011 Paris, France;
          - Annie Sainsard-Chanet, Carole H. Sellem,Centre de Genetique
          Moleculaire, UPR CNRS 9061, batiment 24, allee de la terrasse,
          F-91190 Gif sur Yvette, France;
          DNA sequence of five overlapping  bac clones:
          Bac DP26B06 from 1 to 41045; Bac DP34F04 from 1543 to 105110; Bac
          DP16D11 from 101968 to 174388; Bac DP09G08 from 167044 to 234752;
          Bac DP35C12 from 220939 to 322194.
FEATURES             Location/Qualifiers
     source          1. .322194
                     /organism="Podospora anserina"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5145"
                     /chromosome="5"
                     /clone="DP26B06, DP34F04, DP16D11, DP09G08, DP35C12"
     repeat_region   1347. .1983
                     /note="repeat 2"
     repeat_unit     3393. .3584
                     /note="repeat_region"
     repeat_unit     complement(4218. .4374)
                     /note="repeat_region"
     repeat_unit     complement(5033. .5217)
                     /note="repeat_region"
     misc_feature    6266. .6293
                     /note="palindromic region"
     repeat_unit     6871. .6997
                     /note="repeat_region"
     CDS             complement(7625. .9193)
                     /note="unnamed protein product; hypothetical protein
                     Pa5D0001"
                     /codon_start=1
                     /protein_id="CAD60688.1"
                     /db_xref="GI:27802985"
                     /db_xref="UniProt/TrEMBL:Q875C1"
                     /translation="MVLSSPVSIDPIGSLYTGLLSPQSPNHSQSPLPLRHQQSQADNR
                     STLSDHHNKLDPALLTDRPTFSHDSPIFKRLQHNTDEASSISTPRSVPVVTSAQKGKS
                     KWRPKFTTSKKPPAGVIADSSSLSSTTLEAQKLEEITLTPLCHPETSRERE IVEEYQC
                     YSLAKLYPRSILDPAAHTCVGCCHFPSHHDAGHSAREHLYPGSRRQDSFSICYWNKRS
                     ETYGLHYHLSPFSTQSDDFLQLRIVNLVQPTVSVIEYRIPSTLWAKALAIDREENYVV
                     VGFENATVRFFKAKTSEQPREDRLHASSHRDCRGCPSVDTLAFSQDGFTLLASTRSSK
                     TGLVQIFSWRFPFHEFQELAACRYLIPLHESEDNGLSAAIFRPGSDGEETLVCITTWT
                     QSGTPILIQPQDGQRSEIKVDGSGKSTKLGTRIQCAAFSPSGRELVMIYDRGYVFEVS
                     NLNSSPIDIRRVAISKELTAKSESFAMSFVTIADEENVLVAWADSARATGWIKKIPMT
                     GRVSNNTTLWRGEK"
     CDS             complement(join(9742. .10156,10229. .10924,11059. .11582))
                     /note="unnamed protein product; hypothetical protein
                     Pa5D0002"
                     /codon_start=1
                     /protein_id="CAD60689.1"
                     /db_xref="GI:27802986"
                     /db_xref="UniProt/TrEMBL:Q875C0"
                     /translation="MATDEPNITHGSFCNPSLIRRVVASPDNCSGLASYMFTLCTTVR
                     CGCEPSHLYSTDLHPLVLLVPVHPARPTPSQWRGPRHCSDGIPPPPTKAVNLVLSFAL
                     VAGYRLVLPLPPFSTAYHFPISVPNLPCPSDSLCTSGPTPLGDVDIQMKSENSPIETR
                     LGLIIGNQLGKASENTMPSSLLLRLTNVWRALVLRSDVDDIVLELLGCFSLERMYADA
                     SDRDTYQQLVITLLSQLKSIFDIHRLTLPVVNVHTAWYLGFLVSWEVALRSVEFVLQT
                     INESRETIWENRLLRERYLAGFLLSSLRVLTYARPIFTILPKAGDFGDQVTNRPSIDF
                     TPNPQRSSDLETAVSVLLAFTVSWSRFSIAFLARTRFHSRLARRSLPSSTLTLTAWVF
                     PLHWHPDYQIWHRSCLRSPLSSARCYRRHCASRLPWRLACKISSITRRVAIRSCRLNP
                     ICCQQGDSRTASRASPGAFQKCGSECARCLTYPPPILPGGDGCNTEHYFSYSLTRHPG
                     PISSRHRRDLCRGIRTGCIGCSIREACSPSGYSPRQ"
     CDS             join(12887. .13028,13129. .13291,13370. .13430,13464. .13507,
                     13574. .13673,13997. .14037,14098. .14533)
                     /note="unnamed protein product; part of an hypothetical
                     protein Pa5D0003 similar to a Valine/tyrosine/tryptophan
                     amino-acid permease"
                     /codon_start=1
                     /protein_id="CAD60690.1"
                     /db_xref="GI:27802987"
                     /db_xref="GOA:Q86ZL7"
                     /db_xref="InterPro:IPR002114"
                     /db_xref="InterPro:IPR002293"
                     /db_xref="InterPro:IPR004841"
                     /db_xref="UniProt/TrEMBL:Q86ZL7"
                     /translation="MVCALFLNRIGVEDLLIAGIGLFLGSGRSLASAGPGGAVLCYIL
                     MGTVQNYHLFCVICPDLLTVTSHRVKMTALMSVNAPVMEFPRRFMSRGVGFAVGWVYW
                     FAYAVLAGEQLVAVSNAINSRWVRGLIMPCGYPCFRPMLMLPFKVFGELEYIFGCFKL
                     MFIILLIMVGHSIEGSLGTLLGVWTTFTSVMSSYIGMDIVAATAAESRALSDVESMKM
                     AARKINLRVITLYSLAMLPASFVVPMDHPFINGHSTSVGARSIFIIAVVEAGMPKVAH
                     FFNAVFVFSAFTCGINSMYIATRVLHTLALRGQTGPEFITRRLRQCRSGVPS"
     CDS             complement(join(16250. .16546,16619. .17011,17079. .17408,
                     17498. .17740))
                     /note="unnamed protein product; hypothetical protein
                     Pa5D0004 similar to Dihydrolipoamide succinyltransferase"
                     /codon_start=1
                     /protein_id="CAD60691.1"
                     /db_xref="GI:27802988"
                     /db_xref="GOA:Q86ZL6"
                     /db_xref="InterPro:IPR000089"
                     /db_xref="InterPro:IPR001078"
                     /db_xref="InterPro:IPR003016"
                     /db_xref="InterPro:IPR006255"
                     /db_xref="UniProt/TrEMBL:Q86ZL6"
                     /translation="MVYRGLRLAARAAPKFALYNNNTALRQLPLQFQHVRTYADKIVK
                     VPQMAESISEGTLKQWNKSVGDFVEQDEEIATIETDKIDVAVNAPEAGVIKEFLANEE
                     DTVVVGQDLVRIELGGAPSGDKPAAEPKESAPEKKAEPEKAPEPKQEESKPAASAPPP
                     PKQETPEPKSKPAPAAAPEKPATLGNREEKRVKMNRMRLRIAERLKQSQNTAASLTTF
                     NEVDMSALMEFRNKYKEEVLKKTGVKLGFMSAFSRAAVLAMRDIPAVNASIEGPNGGD
                     TIVYRDYVDISVAVATEKGLVTPVVRNAEAMDLVGIERSIAELGKKARDGKLTIEDMA
                     GGTFTISNGGVFGSLMGTPIINLPQSAVLGLHATKERPVVVNGKIEIRPMMYLALTYD
                     HRLLDGREAVQFLVKVKEYIEDPRKMLL"
     CDS             join(<18816. .18950,19031. .19109,19170. .19420)
                     /note="unnamed protein product; part of an hypothetical
                     protein Pa5D0005"
                     /codon_start=1
                     /protein_id="CAD60692.1"
                     /db_xref="GI:27802989"
                     /db_xref="GOA:Q875B9"
                     /db_xref="InterPro:IPR008011"
                     /db_xref="UniProt/TrEMBL:Q875B9"
                     /translation="RSLYRRSLKLALDWSVHRDVWRGQALYIRSLFEKNRAVTDPRLQ
                     RALLKETEKLLEKWKHPDPYIPPTAPGGSKYERNLQVPNLDRESAALPIILNVQVAWL
                     TYVFVCSSPAAQVLIQHWQWRWNVHKVRRGDWSGEQTHAMGDVECKVVASNL"
     CDS             complement(join(19815. .19873,19929. .20221,20371. .20426))
                     /note="unnamed protein product; hypothetical protein
                     Pa5D0006 similar to histone H2A"
                     /codon_start=1
                     /protein_id="CAD60693.1"
                     /db_xref="GI:27802990"
                     /db_xref="GOA:Q875B8"
                     /db_xref="InterPro:IPR000886"
                     /db_xref="InterPro:IPR002119"
                     /db_xref="InterPro:IPR007124"
                     /db_xref="InterPro:IPR007125"
                     /db_xref="UniProt/TrEMBL:Q875B8"
                     /translation="MTGGGKSGGKASSGKNAQSRSSKAGLAFPVGRVHRLLRKGNYAQ
                     RVGAGAPVYLAAVLEYLAAEILELAGNAARDNKKTRIIPRHLQLAIRNDEELNKLLGH
                     VTIAQGGVLPNIHQNLLPKKTGTKPGKNASQEL"
     CDS             join(22923. .23085,23173. .23244,23312. .23490)
                     /note="unnamed protein product; hypothetical protein
                     Pa5D0007 similar to histone H2B"
                     /codon_start=1
                     /protein_id="CAD60694.1"
                     /db_xref="GI:27802991"
                     /db_xref="GOA:Q875B7"
                     /db_xref="InterPro:IPR000558"
                     /db_xref="InterPro:IPR007124"
                     /db_xref="InterPro:IPR007125"
                     /db_xref="UniProt/TrEMBL:Q875B7"
                     /translation="MPPKAADKKPANKAPATASKAPEKKDAGKKTAASGEKKKRTKAR
                     KETYSSYIYKVLKQVHPDTGISNRAMSILNSFVNDIFERVATEASKLAAYNKKSTISS
                     REIQTSVRLILPGELAKHAVSEGTKAVTKYSSSTK"
     CDS             complement(join(23986. .24472,24529. .26198))
                     /note="unnamed protein product; hypothetical protein
```

```
                     Pa5D0008"
                     /codon_start=1
                     /protein_id="CAD60695.1"
                     /db_xref="GI:27802992"
                     /db_xref="GOA:Q875B6"
                     /db_xref="InterPro:IPR000467"
                     /db_xref="InterPro:IPR001374"
                     /db_xref="UniProt/TrEMBL:Q875B6"
                     /translation="MQSPTLLTFDLPPESRAMSVPFPQRRFSRVSRTQMVRFPDSALA
                     WLQDLAHLTPTDFSLKDEVRNTQSHHLDSAWTSGTVKLRQKPVSFVSAGYSEPLKLLE
                     DIEQDKVPPGEATESKADDMDITTVDVNLNVSAEAVIEKTTLVELDSAAKGKMVLSNS
                     TSQDNPSSETPQEQAQDLFFFDVSGDKAIRDKHRAVHPPPLVPIRKPSLAESDSSEEV
                     ILFRGRAGNAKTVPQSNFVVRNGVATNTTTAITPTGDKTKPTAAEETSLRDDPEVIPV
                     AREKRAGRQRSRSKASKIPKTDEDDEEDAILADYIANMSANPEDDFISDQLRSFNNRR
                     DLGGDNFALNLGSGDENDMPVVEDLSGDEQQAESSGSGLSDADEDNGDENDEADEDDD
                     MDADMDDEGLARLLAKQEELGLGSDELVLIPESFGVSKRGAKKGGQKRATPSSFAKSA
                     NASSVADAFDDLDLADWTVPVPRKRRSKQPPNFNISDSEIEAKLKLDWSRDRERKKER
                     KLARESLRGQGLLDKNASPDDLRVKYPVGLRLEDFKTELVAFLISSDERLEFPPLDKH
                     GRMVLHQLALKFNVKSQSTGKGTTRRPVLYRSKRTITFKPHQITEATRQVDGAARRVG
                     RKYFPRADVAGPRGDTPRDGFRGSAHVSVKALVLREGEVVGASAPELGQENKGRAMLE
                     KMGWSKGMALGALENKGILEPVAQVVKKSKAGLGRT"
     CDS             join(27135. .27879,27949. .28117,28189. .28389,28453. .28840)
                     /note="unnamed protein product; hypothetical protein
                     Pa5D0009 similar to Aut2p Involved in autophagy in
                     Saccharomyces cerevisiae"
                     /codon_start=1
                     /protein_id="CAD60696.1"
                     /db_xref="GI:27802993"
                     /db_xref="InterPro:IPR005078"
                     /db_xref="UniProt/TrEMBL:Q86ZL5"
                     /translation="MKSTRSGSNNWQSAVSGTSPKIATAMEAAISAGAEVSRVGRRIL

Alignment Scores:
Pred. No.:                5.61          Length:          322194
Score:                    84.00         Matches:         12
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       15            Gaps:            0

US-09-858-332G-15 (1-12) x CNS09S4S (1-322194)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||||||
Db     272327 CACAACCACAACCACAACCACAATCACAATCACAAC 272362

RESULT 126
AC131257
LOCUS       AC131257               72409 bp    DNA     linear   HTG 26-SEP-2002
DEFINITION  Homo sapiens chromosome 15 clone RP13-423F3 map 15, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC131257
VERSION     AC131257.3  GI:23322701
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 72409)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP13-423F3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 72409)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 72409)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-SEP-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Sep 26, 2002 this sequence version replaced gi:22657542.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L28022
                Center clone name: 423_F_3
            --------------.
            * NOTE: This record contains 86 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      713: contig of 713 bp in length
            *      714      813: gap of 100 bp
            *      814     1543: contig of 730 bp in length
            *     1544     1643: gap of 100 bp
            *     1644     2398: contig of 755 bp in length
            *     2399     2498: gap of 100 bp
            *     2499     3220: contig of 722 bp in length
            *     3221     3320: gap of 100 bp
            *     3321     4071: contig of 751 bp in length
            *     4072     4171: gap of 100 bp
            *     4172     4928: contig of 757 bp in length
            *     4929     5028: gap of 100 bp
            *     5029     5783: contig of 755 bp in length
            *     5784     5883: gap of 100 bp
            *     5884     6640: contig of 757 bp in length
```

```
*     6641     6740: gap of 100 bp
*     6741     7484: contig of 744 bp in length
*     7485     7584: gap of 100 bp
*     7585     8329: contig of 745 bp in length
*     8330     8429: gap of 100 bp
*     8430     9147: contig of 718 bp in length
*     9148     9247: gap of 100 bp
*     9248     9992: contig of 745 bp in length
*     9993    10092: gap of 100 bp
*    10093    10821: contig of 729 bp in length
*    10822    10921: gap of 100 bp
*    10922    11666: contig of 745 bp in length
*    11667    11766: gap of 100 bp
*    11767    12497: contig of 731 bp in length
*    12498    12597: gap of 100 bp
*    12598    13321: contig of 724 bp in length
*    13322    13421: gap of 100 bp
*    13422    14166: contig of 745 bp in length
*    14167    14266: gap of 100 bp
*    14267    14999: contig of 733 bp in length
*    15000    15099: gap of 100 bp
*    15100    15847: contig of 748 bp in length
*    15848    15947: gap of 100 bp
*    15948    16693: contig of 746 bp in length
*    16694    16793: gap of 100 bp
*    16794    17536: contig of 743 bp in length
*    17537    17636: gap of 100 bp
*    17637    18383: contig of 747 bp in length
*    18384    18483: gap of 100 bp
*    18484    19222: contig of 739 bp in length
*    19223    19322: gap of 100 bp
*    19323    20050: contig of 728 bp in length
*    20051    20150: gap of 100 bp
*    20151    20895: contig of 745 bp in length
*    20896    20995: gap of 100 bp
*    20996    21747: contig of 752 bp in length
*    21748    21847: gap of 100 bp
*    21848    22595: contig of 748 bp in length
*    22596    22695: gap of 100 bp
*    22696    23446: contig of 751 bp in length
*    23447    23546: gap of 100 bp
*    23547    24299: contig of 753 bp in length
*    24300    24399: gap of 100 bp
*    24400    25155: contig of 756 bp in length
*    25156    25255: gap of 100 bp
*    25256    25994: contig of 739 bp in length
*    25995    26094: gap of 100 bp
*    26095    26856: contig of 762 bp in length
*    26857    26956: gap of 100 bp
*    26957    27697: contig of 741 bp in length
*    27698    27797: gap of 100 bp
*    27798    28541: contig of 744 bp in length
*    28542    28641: gap of 100 bp
*    28642    29384: contig of 743 bp in length
*    29385    29484: gap of 100 bp
*    29485    30222: contig of 738 bp in length
*    30223    30322: gap of 100 bp
*    30323    31069: contig of 747 bp in length
*    31070    31169: gap of 100 bp
*    31170    31903: contig of 734 bp in length
*    31904    32003: gap of 100 bp
*    32004    32747: contig of 744 bp in length
*    32748    32847: gap of 100 bp
*    32848    33600: contig of 753 bp in length
*    33601    33700: gap of 100 bp
*    33701    34456: contig of 756 bp in length
*    34457    34556: gap of 100 bp
*    34557    35310: contig of 754 bp in length
*    35311    35410: gap of 100 bp
*    35411    36158: contig of 748 bp in length
*    36159    36258: gap of 100 bp
*    36259    36991: contig of 733 bp in length
*    36992    37091: gap of 100 bp
*    37092    37834: contig of 743 bp in length
*    37835    37934: gap of 100 bp
*    37935    38672: contig of 738 bp in length
*    38673    38772: gap of 100 bp
*    38773    39518: contig of 746 bp in length
*    39519    39618: gap of 100 bp
*    39619    40343: contig of 725 bp in length
*    40344    40443: gap of 100 bp
*    40444    41189: contig of 746 bp in length
*    41190    41289: gap of 100 bp
*    41290    42042: contig of 753 bp in length
*    42043    42142: gap of 100 bp
*    42143    42898: contig of 756 bp in length
*    42899    42998: gap of 100 bp
*    42999    43759: contig of 761 bp in length
*    43760    43859: gap of 100 bp
*    43860    44616: contig of 757 bp in length
*    44617    44716: gap of 100 bp
*    44717    45466: contig of 750 bp in length
*    45467    45566: gap of 100 bp
*    45567    46312: contig of 746 bp in length
*    46313    46412: gap of 100 bp
*    46413    47151: contig of 739 bp in length
*    47152    47251: gap of 100 bp
*    47252    48003: contig of 752 bp in length
*    48004    48103: gap of 100 bp
*    48104    48861: contig of 758 bp in length
*    48862    48961: gap of 100 bp
*    48962    49715: contig of 754 bp in length

Alignment Scores:
Pred. No.:                5.92          Length:          72409
Score:                    79.00         Matches:         11
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    91.67%        Mismatches:      0
Query Match:              94.05%        Indels:          0
DB:                       14            Gaps:            0

US-09-858-332G-15 (1-12) x AC131257 (1-72409)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||:::||||||||||||
Db      10555 CATAATCATAACCATAACCATAGCCATAATCATAAT 10590

RESULT 127
CR380953_03/c
WPCOMMENT
  Sequence split into 10 fragments  LOCUS CR380953 Accession CR380953
     Fragment Name      Begin     End
     CR380953_00            1    110000
     CR380953_01       100001    210000
     CR380953_02       200001    310000
     CR380953_03       300001    410000
     CR380953_04       400001    510000
     CR380953_05       500001    610000
     CR380953_06       600001    710000
     CR380953_07       700001    810000
     CR380953_08       800001    910000
     CR380953_09       900001    992211
Continuation (4 of 10) of CR380953 from base 300001 (CR380953 Candida glabrata strain CBS

Alignment Scores:
Pred. No.:                8.86          Length:          110000
Score:                    79.00         Matches:         11
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    91.67%        Mismatches:      0
Query Match:              94.05%        Indels:          0
DB:                       15            Gaps:            0

US-09-858-332G-15 (1-12) x CR380953_03 (1-110000)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||:::||||||||||||||||||
```

```
Db      30429 CACAACCACAACCACAGCCACAATCACAATCACAAT 30394

RESULT 128
CR382127_10/c
WPCOMMENT
  Sequence split into 23 fragments  LOCUS CR382127 Accession CR382127
     Fragment Name      Begin     End
     CR382127_00            1    110000
     CR382127_01       100001    210000
     CR382127_02       200001    310000
     CR382127_03       300001    410000
     CR382127_04       400001    510000
     CR382127_05       500001    610000
     CR382127_06       600001    710000
     CR382127_07       700001    810000
     CR382127_08       800001    910000
     CR382127_09       900001   1010000
     CR382127_10      1000001   1110000
     CR382127_11      1100001   1210000
     CR382127_12      1200001   1310000
     CR382127_13      1300001   1410000
     CR382127_14      1400001   1510000
     CR382127_15      1500001   1610000
     CR382127_16      1600001   1710000
     CR382127_17      1700001   1810000
     CR382127_18      1800001   1910000
     CR382127_19      1900001   2010000
     CR382127_20      2000001   2110000
     CR382127_21      2100001   2210000
     CR382127_22      2200001   2303261
Continuation (11 of 23) of CR382127 from base 1000001 (CR382127 Yarrowia lipolytica chro

Alignment Scores:
Pred. No.:                8.86          Length:         110000
Score:                    79.00         Matches:        11
Percent Similarity:       100.00%       Conservative:   1
Best Local Similarity:    91.67%        Mismatches:     0
Query Match:              94.05%        Indels:         0
DB:                       15            Gaps:           0

US-09-858-332G-15 (1-12) x CR382127_10 (1-110000)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||||||||:::|||||||||||||||||||||||||
Db     106869 CACAACCACAGCCACAACCACAACCACAACCACAAC 106834

RESULT 129
CR382127_11/c
WPCOMMENT
  Sequence split into 23 fragments  LOCUS CR382127 Accession CR382127
     Fragment Name      Begin     End
     CR382127_00            1    110000
     CR382127_01       100001    210000
     CR382127_02       200001    310000
     CR382127_03       300001    410000
     CR382127_04       400001    510000
     CR382127_05       500001    610000
     CR382127_06       600001    710000
     CR382127_07       700001    810000
     CR382127_08       800001    910000
     CR382127_09       900001   1010000
     CR382127_10      1000001   1110000
     CR382127_11      1100001   1210000
     CR382127_12      1200001   1310000
     CR382127_13      1300001   1410000
     CR382127_14      1400001   1510000
     CR382127_15      1500001   1610000
     CR382127_16      1600001   1710000
     CR382127_17      1700001   1810000
     CR382127_18      1800001   1910000
     CR382127_19      1900001   2010000
     CR382127_20      2000001   2110000
     CR382127_21      2100001   2210000
     CR382127_22      2200001   2303261
Continuation (12 of 23) of CR382127 from base 1100001 (CR382127 Yarrowia lipolytica chro

Alignment Scores:
Pred. No.:                8.86          Length:         110000
Score:                    79.00         Matches:        11
Percent Similarity:       100.00%       Conservative:   1
Best Local Similarity:    91.67%        Mismatches:     0
Query Match:              94.05%        Indels:         0
DB:                       15            Gaps:           0

US-09-858-332G-15 (1-12) x CR382127_11 (1-110000)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||||||||:::|||||||||||||||||||||||||
Db       6869 CACAACCACAGCCACAACCACAACCACAACCACAAC 6834

RESULT 130
AP003400/c
LOCUS       AP003400              135637 bp    DNA     linear   PRI 15-MAR-2003
DEFINITION  Homo sapiens genomic DNA, chromosome 11q, clone:RP11-643G5,
            complete sequence.
ACCESSION   AP003400
VERSION     AP003400.2  GI:15320508
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA
  JOURNAL   Published Only in Database (2001)
REFERENCE   2  (bases 1 to 135637)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-2001) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     On Aug 27, 2001 this sequence version replaced gi:13366104.
FEATURES             Location/Qualifiers
     source          1. .135637
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-643G5"
ORIGIN

Alignment Scores:
Pred. No.:                10.8          Length:         135637
Score:                    79.00         Matches:        11
Percent Similarity:       100.00%       Conservative:   1
Best Local Similarity:    91.67%        Mismatches:     0
Query Match:              94.05%        Indels:         0
DB:                       8             Gaps:           0

US-09-858-332G-15 (1-12) x AP003400 (1-135637)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||||||||||||||||||||:::|||||||||||||
Db      21374 CATAATCATAACCATAACCATAGCCATAATCATAAT 21339

RESULT 131
AC163335
LOCUS       AC163335              164631 bp    DNA     linear   HTG 03-AUG-2005
DEFINITION  Mus musculus chromosome 7 clone RP24-149G13 map 7, *** SEQUENCING
```

```
            IN PROGRESS ***, 2 ordered pieces.
ACCESSION   AC163335
VERSION     AC163335.5  GI:71725519
KEYWORDS    HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 164631)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 7, clone RP24-149G13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 164631)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
            Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J.,
            Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
            DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
            Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,
            Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,
            Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
            MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
            Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
            Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUN-2005) Broad Institute of MIT and Harvard, 320
            Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 164631)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
            Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J.,
            Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
            DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
            Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,
            Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,
            Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
            MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
            Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
            Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-2005) Broad Institute of MIT and Harvard, 320
            Charles Street, Cambridge, MA 02141, USA
COMMENT     On Aug 3, 2005 this sequence version replaced gi:70909418.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Broad Institute of MIT and Harvard
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@broad.mit.edu
            -------------- Project Information
                Center project name: L33061
                Center clone name: 149_G_13
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1    99799: contig of 99799 bp in length
            *   99800    99899: gap of unknown length
            *   99900   164631: contig of 64732 bp in length.
FEATURES            Location/Qualifiers
     source         1. .164631
                    /organism="Mus musculus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10090"
                    /chromosome="7"
                    /map="7"
                    /clone="RP24-149G13"
                    /clone_lib="RPCI-24 Male Mouse BAC"
     gap            99800..99899
                    /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:                13.1          Length:          164631
Score:                    79.00         Matches:         11
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    91.67%        Mismatches:      0
Query Match:              94.05%        Indels:          0
DB:                       14            Gaps:            0

US-09-858-332G-15 (1-12) x AC163335 (1-164631)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||:::||||||
Db     113857 CACAACCACAACCACAACCACAACCACCACCACAAC 113892

RESULT 132
AC012198/c
LOCUS       AC012198              168425 bp    DNA     linear   HTG 26-MAR-2000
DEFINITION  Homo sapiens chromosome 15 clone RP11-345O16 map 15, WORKING DRAFT
            SEQUENCE, 18 unordered pieces.
ACCESSION   AC012198
VERSION     AC012198.3  GI:7329279
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 168425)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-345O16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 168425)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
```

```
  TITLE     Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  JOURNAL   Direct Submission
            Submitted (21-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 26, 2000 this sequence version replaced gi:6479156.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L1687
                Center clone name: 345_O_16
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 158972 bases at least Q40
                Consensus quality: 163604 bases at least Q30
                Consensus quality: 165235 bases at least Q20
                Insert size: 182000; agarose-fp
                Insert size: 166725; sum-of-contigs
                Quality coverage: 4.2 in Q20 bases; agarose-fp
                Quality coverage: 4.6 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 18 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1003: contig of 1003 bp in length
            *     1004     1103: gap of 100 bp
            *     1104     2522: contig of 1419 bp in length
            *     2523     2622: gap of 100 bp
            *     2623     4412: contig of 1790 bp in length
            *     4413     4512: gap of 100 bp
            *     4513     6455: contig of 1943 bp in length
            *     6456     6555: gap of 100 bp
            *     6556     8986: contig of 2431 bp in length
            *     8987     9086: gap of 100 bp
            *     9087    11922: contig of 2836 bp in length
            *    11923    12022: gap of 100 bp
            *    12023    15346: contig of 3324 bp in length
            *    15347    15446: gap of 100 bp
            *    15447    20136: contig of 4690 bp in length
            *    20137    20236: gap of 100 bp
            *    20237    26216: contig of 5980 bp in length
            *    26217    26316: gap of 100 bp
            *    26317    36347: contig of 10031 bp in length
            *    36348    36447: gap of 100 bp
            *    36448    45053: contig of 8606 bp in length
            *    45054    45153: gap of 100 bp
            *    45154    55860: contig of 10707 bp in length
            *    55861    55960: gap of 100 bp
            *    55961    67634: contig of 11674 bp in length
            *    67635    67734: gap of 100 bp
            *    67735    80652: contig of 12918 bp in length
            *    80653    80752: gap of 100 bp
            *    80753    99292: contig of 18540 bp in length
            *    99293    99392: gap of 100 bp
            *    99393   115499: contig of 16107 bp in length
            *   115500   115599: gap of 100 bp
            *   115600   134093: contig of 18494 bp in length
            *   134094   134193: gap of 100 bp
            *   134194   168425: contig of 34232 bp in length.
FEATURES            Location/Qualifiers
     source         1. .168425
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /chromosome="15"
                    /map="15"
                    /clone="RP11-345O16"
                    /clone_lib="RPCI-11 Human Male BAC"
     misc_feature   1. .1003
                    /note="assembly_fragment"
     gap            1004. .1103
                    /estimated_length=100
     misc_feature   1104. .2522
                    /note="assembly_fragment"
     gap            2523. .2622
                    /estimated_length=100
     misc_feature   2623. .4412
                    /note="assembly_fragment"
     gap            4413. .4512
                    /estimated_length=100
     misc_feature   4513. .6455
                    /note="assembly_fragment"
     gap            6456. .6555
                    /estimated_length=100
     misc_feature   6556. .8986
                    /note="assembly_fragment"
     gap            8987. .9086
                    /estimated_length=100
     misc_feature   9087. .11922
                    /note="assembly_fragment"
     gap            11923. .12022
                    /estimated_length=100
     misc_feature   12023. .15346
                    /note="assembly_fragment"
     gap            15347. .15446
                    /estimated_length=100
     misc_feature   15447. .20136
                    /note="assembly_fragment"
     gap            20137. .20236
                    /estimated_length=100
     misc_feature   20237. .26216
                    /note="assembly_fragment"
     gap            26217. .26316
                    /estimated_length=100
     misc_feature   26317. .36347
                    /note="assembly_fragment"
     gap            36348. .36447
                    /estimated_length=100
     misc_feature   36448. .45053
                    /note="assembly_fragment
                    clone_end:T7
                    vector_side:left"
     gap            45054. .45153
                    /estimated_length=100
     misc_feature   45154. .55860
                    /note="assembly_fragment"
     gap            55861. .55960
                    /estimated_length=100
     misc_feature   55961. .67634
                    /note="assembly_fragment"
     gap            67635. .67734
                    /estimated_length=100
     misc_feature   67735. .80652
                    /note="assembly_fragment"
     gap            80653. .80752
                    /estimated_length=100
     misc_feature   80753. .99292
                    /note="assembly_fragment"
     gap            99293. .99392
                    /estimated_length=100
     misc_feature   99393. .115499
                    /note="assembly_fragment
                    clone_end:SP6
                    vector_side:left"
```

```
     gap             115500. .115599
                     /estimated_length=100
     misc_feature    115600. .134093
                     /note="assembly_fragment"
     gap             134094. .134193
                     /estimated_length=100
     misc_feature    134194. .168425
                     /note="assembly_fragment"
ORIGIN

Alignment Scores:
Pred. No.:              13.4          Length:         168425
Score:                  79.00         Matches:        11
Percent Similarity:     100.00%       Conservative:   1
Best Local Similarity:  91.67%        Mismatches:     0
Query Match:            94.05%        Indels:         0
DB:                     14            Gaps:           0

US-09-858-332G-15 (1-12) x AC012198 (1-168425)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||:::||||||||||||
Db      51561 CATAATCATAACCATAACCATAGCCATAATCATAAT 51526

RESULT 133
AC132952/c
LOCUS       AC132952               177345 bp    DNA     linear   ROD 26-JAN-2005
DEFINITION  Mus musculus BAC clone RP24-175H8 from 19, complete sequence.
ACCESSION   AC132952
VERSION     AC132952.3  GI:50345242
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 177345)
  AUTHORS   Trani,L., Kozlowicz,A., Haglund,K., Bielicki,L. and Meyer,R.
  TITLE     The sequence of Mus musculus BAC clone RP24-175H8
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 177345)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-SEP-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 177345)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 177345)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2004) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   5  (bases 1 to 177345)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JAN-2005) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jul 16, 2004 this sequence version replaced gi:32441371.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: M_BB0175H08
            --------------.

            NOTICE:
                This sequence was finished as follows unless otherwise noted:
                all regions were double stranded, sequenced with an alternate
                chemistry, or covered by high quality data (i.e., phred quality >=
                30); an attempt was made to resolve all sequencing problems, such
                as compressions and repeats; all regions were covered by sequence
                from more than one subclone; and the assembly was confirmed by
                restriction digest.

                MAPPING INFORMATION:
                Mapping information for this clone was provided by Dr. Wes Warren,
                Department of Genetics, Washington University, St. Louis MO. For
                additional information about the map position of this sequence, see
                http://genome.wustl.edu

                SOURCE INFORMATION:
                The RPCI-24 BAC Library has been constructed by Pieter de Jong and
                coworkers (http://www.chori.org) from male C57BL/6J mouse spleen
                and/or brain genomic DNA. The clone and detailed information can be
                obtained from Pieter de Jong and coworkers at http://www.chori.org

                NEIGHBORING SEQUENCE INFORMATION:
                This sequence is the entire insert of the clone. This clone is
                overlapped by AC147229.
FEATURES            Location/Qualifiers
     source         1. .177345
                    /organism="Mus musculus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10090"
                    /chromosome="19"
                    /map="19"
                    /clone="RP24-175H8"
                    /clone_lib="RPCI-24"
     repeat_region  277. .903
                    /rpt_family="L1"
     repeat_region  902. .1533
                    /rpt_family="L1"
     repeat_region  1681. .2064
                    /rpt_family="MaLR"
     repeat_region  2505. .2868
                    /rpt_family="L1"
     repeat_region  2875. .3050
                    /rpt_family="B2"
     repeat_region  2920. .3063
                    /rpt_family="B4"
     repeat_region  3064. .3132
                    /rpt_family="Alu"
     repeat_region  3286. .3394
                    /rpt_family="Alu"
     repeat_region  3522. .3787
                    /rpt_family="L1"
     repeat_region  3990. .4117
                    /rpt_family="ERVL"
     repeat_region  4301. .4543
                    /rpt_family="B4"
     repeat_region  4546. .4663
                    /rpt_family="ERVL"
     repeat_region  4770. .5184
                    /rpt_family="L1"
     repeat_region  5264. .5449
                    /rpt_family="B2"
     repeat_region  5477. .6207
                    /rpt_family="L1"
     repeat_region  6208. .6295
                    /rpt_family="B2"
     repeat_region  6847. .7050
                    /rpt_family="L1"
     repeat_region  7191. .7283
                    /rpt_family="L1"
     repeat_region  7634. .7757
                    /rpt_family="B4"
     repeat_region  10715. .10908
                    /rpt_family="MER1_type"
     repeat_region  11541. .11888
```

```
                     /rpt_family="L1"
repeat_region        11949. .12796
                     /rpt_family="L1"
repeat_region        15322. .15385
                     /rpt_family="ERV1"
repeat_region        15975. .16264
                     /rpt_family="MaLR"
repeat_region        17548. .17774
                     /rpt_family="L1"
repeat_region        17777. .17972
                     /rpt_family="B2"
repeat_region        18017. .18219
                     /rpt_family="L1"
repeat_region        18283. .18407
                     /rpt_family="Alu"
repeat_region        18414. .19344
                     /rpt_family="L1"
repeat_region        19407. .19553
                     /rpt_family="L1"
repeat_region        19720. .20046
                     /rpt_family="MaLR"
repeat_region        20392. .20542
                     /rpt_family="MaLR"
repeat_region        20547. .20702
                     /rpt_family="L1"
repeat_region        20911. .21144
                     /rpt_family="MER1_type"
repeat_region        21294. .21460
                     /rpt_family="MER1_type"
repeat_region        21586. .21990
                     /rpt_family="MaLR"
repeat_region        22029. .22115
                     /rpt_family="MER1_type"
repeat_region        22116. .22227
                     /rpt_family="Alu"
repeat_region        22228. .22847
                     /rpt_family="MER1_type"
repeat_region        22874. .23332
                     /rpt_family="L1"
repeat_region        23705. .23898
                     /rpt_family="MaLR"
repeat_region        24917. .25033
                     /rpt_family="MaLR"
repeat_region        25722. .25866
                     /rpt_family="Alu"
repeat_region        28096. .28260
                     /rpt_family="B4"
repeat_region        28779. .29123
                     /rpt_family="ERVK"
repeat_region        29610. .29750
                     /rpt_family="Alu"
repeat_region        30530. .30845
                     /rpt_family="MaLR"
repeat_region        32467. .32846
                     /rpt_family="L1"
repeat_region        32863. .33858
                     /rpt_family="L1"
repeat_region        34302. .34389
                     /rpt_family="L1"
repeat_region        34490. .34665
                     /rpt_family="L1"
repeat_region        34710. .34908
                     /rpt_family="MaLR"
repeat_region        34910. .36291
                     /rpt_family="L1"
repeat_region        36286. .36895
                     /rpt_family="L1"
repeat_region        36893. .37148
                     /rpt_family="L1"
repeat_region        37641. .37933
                     /rpt_family="L1"
repeat_region        37933. .38665
                     /rpt_family="L1"
repeat_region        38768. .38906
                     /rpt_family="B4"
unsure               43598. .47320
                     /note="Unresolved duplication."
repeat_region        49344. .49423
                     /rpt_family="B4"
repeat_region        50601. .51106

Alignment Scores:
Pred. No.:                14              Length:          177345
Score:                    79.00           Matches:         11
Percent Similarity:       100.00%         Conservative:    1
Best Local Similarity:    91.67%          Mismatches:      0
Query Match:              94.05%          Indels:          0
DB:                       9               Gaps:            0

US-09-858-332G-15 (1-12) x AC132952 (1-177345)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||:::
Db     160627 CACAACCACAACCACAACCACAACCACAACCACCAC 160592

RESULT 134
AP001927/c
LOCUS       AP001927               188256 bp    DNA     linear   HTG 25-AUG-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-482J6 map 11q, WORKING DRAFT
            SEQUENCE, 7 unordered pieces.
ACCESSION   AP001927
VERSION     AP001927.3  GI:9927287
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 188256)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 188,256 genomic DNA of 11q
  JOURNAL   Published Only in DataBase (2000)
REFERENCE   2  (bases 1 to 188256)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-APR-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-42-778-9923, Fax:81-42-778-9924)
COMMENT     On Aug 26, 2000 this sequence version replaced gi:8117574.
            ---------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.riken.go.jp/
                Contact: hattori@gsc.riken.go.jp
            ---------- Project Information
                Center project name: HumDraft11
                Center clone name: RP11-482J6
            ---------- Summary Statistics
                Sequencing vector: PCR products; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 186386 bases at least Q40
                Consensus quality: 187078 bases at least Q30
                Consensus quality: 187448 bases at least Q20
                Insert size: 187656; sum-of-contigs
                Quality coverage: 11.55x in Q20 bases; sum-of-contigs
            ----------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            7 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
```

```
            as soon as it is available and the accession number will be
            preserved
                   1     71392 contig of  71392 bp in length
                71493    107532 contig of  36040 bp in length
               107633    135854 contig of  28222 bp in length
               135955    159520 contig of  23566 bp in length
               159621    173994 contig of  14374 bp in length
               174095    183415 contig of   9321 bp in length
               183516    188256 contig of   4741 bp in length.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    71392: contig of 71392 bp in length
            *   71393    71492: gap of 100 bp
            *   71493   107532: contig of 36040 bp in length
            *  107533   107632: gap of 100 bp
            *  107633   135854: contig of 28222 bp in length
            *  135855   135954: gap of 100 bp
            *  135955   159520: contig of 23566 bp in length
            *  159521   159620: gap of 100 bp
            *  159621   173994: contig of 14374 bp in length
            *  173995   174094: gap of 100 bp
            *  174095   183415: contig of 9321 bp in length
            *  183416   183515: gap of 100 bp
            *  183516   188256: contig of 4741 bp in length.
FEATURES             Location/Qualifiers
     source          1. .188256
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-482J6"
     misc_feature    1. .71392
                     /note="assembly_fragment clone_end:T7 vector_side:left"
     misc_feature    71493. .107532
                     /note="assembly_fragment"
     misc_feature    107633. .135854
                     /note="assembly_fragment"
     misc_feature    135955. .159520
                     /note="assembly_fragment"
     misc_feature    159621. .173994
                     /note="assembly_fragment clone_end:SP6 vector_side:left"
     misc_feature    174095. .183415
                     /note="assembly_fragment"
     misc_feature    183516. .188256
                     /note="assembly_fragment"
ORIGIN

Alignment Scores:
Pred. No.:              14.9          Length:        188256
Score:                  79.00         Matches:       11
Percent Similarity:     100.00%       Conservative:  1
Best Local Similarity:  91.67%        Mismatches:    0
Query Match:            94.05%        Indels:        0
DB:                     14            Gaps:          0

US-09-858-332G-15 (1-12) x AP001927 (1-188256)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||:::||||||||||||
Db     184134 CATAATCATAACCATAACCATAGCCATAATCATAAT 184099

RESULT 135
AC151575
LOCUS       AC151575              200563 bp    DNA     linear   ROD 26-JAN-2005
DEFINITION  Mus musculus BAC clone RP24-126E1 from 19, complete sequence.
ACCESSION   AC151575
VERSION     AC151575.1  GI:52077906
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 200563)
  AUTHORS   Kruchowski,S., Haglund,K. and Haakenson,W.
  TITLE     The sequence of Mus musculus BAC clone RP24-126E1
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 200563)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-2004) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 200563)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JAN-2005) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: M_BB0126E01
            --------------.

            NOTICE:

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-24 BAC Library has been constructed by Pieter de Jong and
            coworkers (http://www.chori.org) from male C57BL/6J mouse spleen
            and/or brain genomic DNA. The clone and detailed information can be
            obtained from Pieter de Jong and coworkers at http://www.chori.org

            NEIGHBORING SEQUENCE INFORMATION:
            This sequence is the entire insert of the clone.
FEATURES             Location/Qualifiers
     source          1. .200563
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="19"
                     /map="19"
                     /clone="RP24-126E1"
                     /clone_lib="RPCI-24"
     unsure          11021. .11146
                     /note="Sequence derived from one plasmid subclone."
     unsure          42485. .42555
                     /note="Sequence derived from one plasmid subclone."
     unsure          191089. .191206
                     /note="Unresolved simple sequence repeat."
ORIGIN

Alignment Scores:
```

```
Pred. No.:              15.8          Length:        200563
Score:                  79.00         Matches:       11
Percent Similarity:     100.00%       Conservative:  1
Best Local Similarity:  91.67%        Mismatches:    0
Query Match:            94.05%        Indels:        0
DB:                     9             Gaps:          0

US-09-858-332G-15 (1-12) x AC151575 (1-200563)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||:::
Db     114027 CACAACCACAACCACAACCACAACCACAACCACCAC 114062

RESULT 136
AC166819
LOCUS       AC166819              202116 bp    DNA     linear   HTG 09-AUG-2005
DEFINITION  Mus musculus chromosome 6 clone RP24-492L15, *** SEQUENCING IN
            PROGRESS ***, 7 unordered pieces.
ACCESSION   AC166819
VERSION     AC166819.1  GI:71980214
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 202116)
  AUTHORS   Wilson,R.K.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 202116)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-2005) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT
            ----------------- Genome Center ----------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu
            -------------- Project Information ------------------
            Center project name: M_BB0492L15
            ----------------------------------------------------
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 198086 bases at least Q40
            Consensus quality: 199095 bases at least Q30
            Consensus quality: 199855 bases at least Q20
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1535: contig of 1535 bp in length
            *     1536     1635: gap of unknown length
            *     1636    10240: contig of 8605 bp in length
            *    10241    10340: gap of unknown length
            *    10341    24457: contig of 14117 bp in length
            *    24458    24557: gap of unknown length
            *    24558    37426: contig of 12869 bp in length
            *    37427    37526: gap of unknown length
            *    37527    57241: contig of 19715 bp in length
            *    57242    57341: gap of unknown length
            *    57342   199980: contig of 142639 bp in length
            *   199981   200080: gap of unknown length
            *   200081   202116: contig of 2036 bp in length.
FEATURES            Location/Qualifiers
     source         1. .202116
                    /organism="Mus musculus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10090"
                    /chromosome="6"
                    /clone="RP24-492L15"
     misc_feature   1. .1535
                    /note="assembly_name:Contig17"
     gap            1536. .1635
                    /estimated_length=unknown
     misc_feature   1636. .10240
                    /note="assembly_name:Contig18"
     gap            10241. .10340
                    /estimated_length=unknown
     misc_feature   10341. .24457
                    /note="assembly_name:Contig19"
     gap            24458. .24557
                    /estimated_length=unknown
     misc_feature   24558. .37426
                    /note="assembly_name:Contig20
                    clone_end:T7
                    vector_side:left"
     gap            37427. .37526
                    /estimated_length=unknown
     misc_feature   37527. .57241
                    /note="assembly_name:Contig21
                    clone_end:SP6
                    vector_side:right"
     gap            57242. .57341
                    /estimated_length=unknown
     misc_feature   57342. .199980
                    /note="assembly_name:Contig22"
     gap            199981. .200080
                    /estimated_length=unknown
     misc_feature   200081. .202116
                    /note="assembly_name:Contig9"
ORIGIN

Alignment Scores:
Pred. No.:              15.9          Length:        202116
Score:                  79.00         Matches:       11
Percent Similarity:     100.00%       Conservative:  1
Best Local Similarity:  91.67%        Mismatches:    0
Query Match:            94.05%        Indels:        0
DB:                     14            Gaps:          0

US-09-858-332G-15 (1-12) x AC166819 (1-202116)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||:::|||||||||||||||||||
Db     165020 CACAATCACAATCACAGTCACAATCACAACCACAAT 165055

RESULT 137
AC127100
LOCUS       AC127100              222745 bp    DNA     linear   HTG 11-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-164M9, *** SEQUENCING IN PROGRESS
            ***, 2 unordered pieces.
ACCESSION   AC127100
VERSION     AC127100.2  GI:23664869
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 222745)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
```

Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 222745)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (13-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 222745)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (11-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Oct 10, 2002 this sequence version replaced gi:21743892.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GNPW
Center clone name: CH230-164M9
---------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 198734 bases at least Q40
Consensus quality: 200599 bases at least Q30
Consensus quality: 201828 bases at least Q20
Estimated insert size: 205186; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 221310: contig of 221310 bp in length
* 221311 221410: gap of unknown length
* 221411 222745: contig of 1335 bp in length.

FEATURES Location/Qualifiers
source 1. .222745
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-164M9"
misc_feature 1. .1603
/note="wgs_end_extension
clone_end:T7"
misc_feature 3684. .4529
/note="clone_boundary
clone_end:T7
site:EcoRI
end_sequence:BH362192"
gap 221311. .221410
/estimated_length=unknown

ORIGIN

Alignment Scores:

Pred. No.:	17.5	Length:	222745
Score:	79.00	Matches:	11
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	91.67%	Mismatches:	0
Query Match:	94.05%	Indels:	0
DB:	14	Gaps:	0

US-09-858-332G-15 (1-12) x AC127100 (1-222745)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||:::
Db     115853 CACAACCACAACCACAACCACAACCACAACCACAGC 115888
```

RESULT 138
AC112891/c
LOCUS AC112891 260910 bp DNA linear HTG 20-NOV-2002
DEFINITION Rattus norvegicus clone CH230-74K14, WORKING DRAFT SEQUENCE, 2 unordered pieces.
ACCESSION AC112891
VERSION AC112891.5 GI:25138603
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 260910)

AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 260910)

AUTHORS Worley,K.C.

TITLE Direct Submission

JOURNAL Submitted (25-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 260910)

AUTHORS Rat Genome Sequencing Consortium.

TITLE Direct Submission

JOURNAL Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On Nov 20, 2002 this sequence version replaced gi:23321801.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

```
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GSJF
                Center clone name: CH230-74K14
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 240270 bases at least Q40
                Consensus quality: 242948 bases at least Q30
                Consensus quality: 244433 bases at least Q20
                Estimated insert size: 246119; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   259195: contig of 259195 bp in length
            *  259196   259295: gap of unknown length
            *  259296   260910: contig of 1615 bp in length.
FEATURES             Location/Qualifiers
     source          1. .260910
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-74K14"
     misc_feature    1. .1145
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    9348. .10129
                     /note="clone_boundary
                     clone_end:Sp6
                     site:
                     end_sequence:BH338715"
     misc_feature    255023. .256206
                     /note="wgs_contig"
     misc_feature    257856. .259195
                     /note="wgs_contig"
     gap             259196. .259295
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:                20.4          Length:          260910
Score:                    79.00         Matches:         11
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    91.67%        Mismatches:      0
Query Match:              94.05%        Indels:          0
DB:                       14            Gaps:            0

US-09-858-332G-15 (1-12) x AC112891 (1-260910)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||:::
Db      39725 CACAACCACAACCACAACCACAACCACAACCACAGC 39690

RESULT 139
AC145086
LOCUS       AC145086               323443 bp    DNA     linear   HTG 05-JUN-2003
DEFINITION  Mus musculus clone RP24-501D2, WORKING DRAFT SEQUENCE, 25 unordered
            pieces.
```

```
ACCESSION   AC145086
VERSION     AC145086.1  GI:31416088
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 323443)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-501D2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 323443)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUN-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L17867
                Center clone name: 501_D_2
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 25 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1353: contig of 1353 bp in length
            *    1354     1453: gap of 100 bp
            *    1454     1701: contig of 248 bp in length
            *    1702     1801: gap of 100 bp
            *    1802     2474: contig of 673 bp in length
            *    2475     2574: gap of 100 bp
            *    2575     3188: contig of 614 bp in length
            *    3189     3288: gap of 100 bp
            *    3289     4882: contig of 1594 bp in length
            *    4883     4982: gap of 100 bp
            *    4983     6782: contig of 1800 bp in length
            *    6783     6882: gap of 100 bp
            *    6883     8513: contig of 1631 bp in length
            *    8514     8613: gap of 100 bp
            *    8614    10358: contig of 1745 bp in length
            *   10359    10458: gap of 100 bp
            *   10459    12731: contig of 2273 bp in length
            *   12732    12831: gap of 100 bp
            *   12832    14918: contig of 2087 bp in length
            *   14919    15018: gap of 100 bp
            *   15019    16895: contig of 1877 bp in length
            *   16896    16995: gap of 100 bp
            *   16996    21416: contig of 4421 bp in length
            *   21417    21516: gap of 100 bp
            *   21517    26613: contig of 5097 bp in length
            *   26614    26713: gap of 100 bp
            *   26714    85797: contig of 59084 bp in length
            *   85798    85897: gap of 100 bp
            *   85898    93199: contig of 7302 bp in length
            *   93200    93299: gap of 100 bp
            *   93300   106460: contig of 13161 bp in length
            *  106461   106560: gap of 100 bp
            *  106561   120175: contig of 13615 bp in length
            *  120176   120275: gap of 100 bp
            *  120276   131307: contig of 11032 bp in length
            *  131308   131407: gap of 100 bp
            *  131408   149073: contig of 17666 bp in length
            *  149074   149173: gap of 100 bp
            *  149174   166679: contig of 17506 bp in length
            *  166680   166779: gap of 100 bp
            *  166780   185847: contig of 19068 bp in length
            *  185848   185947: gap of 100 bp
            *  185948   210214: contig of 24267 bp in length
            *  210215   210314: gap of 100 bp
            *  210315   254670: contig of 44356 bp in length
            *  254671   254770: gap of 100 bp
            *  254771   318269: contig of 63499 bp in length
            *  318270   318369: gap of 100 bp
            *  318370   323443: contig of 5074 bp in length.
FEATURES             Location/Qualifiers
     source          1. .323443
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP24-501D2"
                     /clone_lib="RPCI-24 Male Mouse BAC"
     misc_feature    1. .1353
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     gap             1354. .1453
                     /estimated_length=100
     misc_feature    1454. .1701
                     /note="assembly_fragment"
     gap             1702. .1801
                     /estimated_length=100
     misc_feature    1802. .2474
                     /note="assembly_fragment"
     gap             2475. .2574
                     /estimated_length=100
     misc_feature    2575. .3188
                     /note="assembly_fragment"
     gap             3189. .3288
                     /estimated_length=100
     misc_feature    3289. .4882
                     /note="assembly_fragment"
     gap             4883. .4982
                     /estimated_length=100
     misc_feature    4983. .6782
                     /note="assembly_fragment"
     gap             6783. .6882
                     /estimated_length=100
     misc_feature    6883. .8513
                     /note="assembly_fragment"
     gap             8514. .8613
                     /estimated_length=100
     misc_feature    8614. .10358
                     /note="assembly_fragment"
     gap             10359. .10458
                     /estimated_length=100
```

```
     misc_feature    10459. .12731
                     /note="assembly_fragment"
     gap             12732. .12831
                     /estimated_length=100
     misc_feature    12832. .14918
                     /note="assembly_fragment"
     gap             14919. .15018
                     /estimated_length=100
     misc_feature    15019. .16895
                     /note="assembly_fragment"
     gap             16896. .16995
                     /estimated_length=100
     misc_feature    16996. .21416
                     /note="assembly_fragment"
     gap             21417. .21516
                     /estimated_length=100
     misc_feature    21517. .26613
                     /note="assembly_fragment"
     gap             26614. .26713
                     /estimated_length=100
     misc_feature    26714. .85797
                     /note="assembly_fragment"
     gap             85798. .85897
                     /estimated_length=100
     misc_feature    85898. .93199
                     /note="assembly_fragment"
     gap             93200. .93299
                     /estimated_length=100
     misc_feature    93300. .106460
                     /note="assembly_fragment"
     gap             106461. .106560
                     /estimated_length=100
     misc_feature    106561. .120175
                     /note="assembly_fragment"
     gap             120176. .120275
                     /estimated_length=100
     misc_feature    120276. .131307
                     /note="assembly_fragment"
     gap             131308. .131407
                     /estimated_length=100
     misc_feature    131408. .149073
                     /note="assembly_fragment"
     gap             149074. .149173
                     /estimated_length=100
     misc_feature    149174. .166679
                     /note="assembly_fragment"
     gap             166680. .166779
                     /estimated_length=100
     misc_feature    166780. .185847
                     /note="assembly_fragment"
     gap             185848. .185947

Alignment Scores:
Pred. No.:              25              Length:         323443
Score:                  79.00           Matches:        11
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  91.67%          Mismatches:     0
Query Match:            94.05%          Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-15 (1-12) x AC145086 (1-323443)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||:::|||||||
Db      29594 CACAACCACAACCACAACCACAACCACCACCACAAC 29629

RESULT 140
AR568831/c
LOCUS       AR568831                 139 bp    DNA     linear   PAT 14-DEC-2004
DEFINITION  Sequence 5 from patent US 6740747.
ACCESSION   AR568831
VERSION     AR568831.1  GI:56568473
KEYWORDS    .                        .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 139)
  AUTHORS   Kaushik,A. and Saini,S.S.
  TITLE     Bovine VDJ cassette, BF1H1, suitable for antigenization
  JOURNAL   Patent: US 6740747-A 5 25-MAY-2004;
FEATURES             Location/Qualifiers
     source          1. .139
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              0.0193          Length:         139
Score:                  78.00           Matches:        11
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  91.67%          Mismatches:     0
Query Match:            92.86%          Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-15 (1-12) x AR568831 (1-139)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              :::||||||||||||||||||||||||||||||||
Db         56 TACAACCATAACCATAACCACAACCATAACCATAAC 21

RESULT 141
AY559839/c
LOCUS       AY559839                 205 bp    DNA     linear   MAM 19-MAR-2004
DEFINITION  Bos taurus germline immunoglobulin heavy chain gene, DH2 segment,
            partial sequence.
ACCESSION   AY559839
VERSION     AY559839.1  GI:45551581
KEYWORDS    .
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 205)
  AUTHORS   Shojaei,F., Saini,S.S. and Kaushik,A.K.
  TITLE     Unusually long germline DH genes contribute to large sized CDR3H in
            bovine antibodies
  JOURNAL   Mol. Immunol. 40 (1), 61-67 (2003)
   PUBMED   12909131
REFERENCE   2  (bases 1 to 205)
  AUTHORS   Shojaei,F., Saini,S.S. and Kaushik,A.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-FEB-2004) Pathobiology, University of Guelph,
            University of Guelph, Guelph, ON N1G 2W1, Canada
FEATURES             Location/Qualifiers
     source          1. .205
                     /organism="Bos taurus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9913"
                     /chromosome="21"
                     /map="21q24"
                     /germline
     gene            <1. .>205
                     /gene="immunoglobulin heavy chain"
     D_segment       1. .205
                     /gene="immunoglobulin heavy chain"
                     /note="DH2"
     repeat_region   1. .9
                     /note="recombination signal sequence"
     repeat_region   23. .29
                     /note="recombination signal sequence"
     repeat_region   177. .184
                     /note="recombination signal sequence"
     repeat_region   197. .205
                     /note="recombination signal sequence"
ORIGIN
```

```
Alignment Scores:
Pred. No.:                0.0281          Length:         205
Score:                    78.00           Matches:        11
Percent Similarity:       100.00%         Conservative:   1
Best Local Similarity:    91.67%          Mismatches:     0
Query Match:              92.86%          Indels:         0
DB:                       4               Gaps:           0

US-09-858-332G-15 (1-12) x AY559839 (1-205)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              :::||||||||||||||||||||||||||||||||||
Db         86 TACAACCATAACCATAACCACAACCATAACCATAAC 51

RESULT 142
AR568828/c
LOCUS       AR568828                205 bp    DNA     linear   PAT 14-DEC-2004
DEFINITION  Sequence 2 from patent US 6740747.
ACCESSION   AR568828
VERSION     AR568828.1  GI:56568470
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 205)
  AUTHORS   Kaushik,A. and Saini,S.S.
  TITLE     Bovine VDJ cassette, BF1H1, suitable for antigenization
  JOURNAL   Patent: US 6740747-A 2 25-MAY-2004;
FEATURES             Location/Qualifiers
     source          1. .205
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:                0.0281          Length:         205
Score:                    78.00           Matches:        11
Percent Similarity:       100.00%         Conservative:   1
Best Local Similarity:    91.67%          Mismatches:     0
Query Match:              92.86%          Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-15 (1-12) x AR568828 (1-205)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              :::||||||||||||||||||||||||||||||||||
Db         86 TACAACCATAACCATAACCACAACCATAACCATAAC 51

RESULT 143
BTA512977/c
LOCUS       BTA512977               211 bp    DNA     linear   MAM 28-JUL-2004
DEFINITION  Bos taurus partial IGHV gene for immunoglobulin heavy chain
            variable region, exon 1 isolate fetus 1, clone 1.
ACCESSION   AJ512977
VERSION     AJ512977.1  GI:24459872
KEYWORDS    IGHV gene; immunoglobulin heavy chain; variable region.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1
  AUTHORS   Panelli,S., Damiani,G., Galli,C. and Sgaramella,V.
  TITLE     Rearranged genomes of bovine blood cells can allow the development
            of clones till late fetal stages; but rare unrearranged genomes
            have greater potential and lead to adulthood
  JOURNAL   Gene 334, 99-103 (2004)
   PUBMED   15256259
REFERENCE   2  (bases 1 to 211)
  AUTHORS   Panelli,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT-2002) Panelli S., Patol. Anim. Igiene e Sanita
            Pubbl. Vet., Fondazione PTP c/o Universita di Milano, via Celoria
            10, Milano 20133, ITALY
FEATURES             Location/Qualifiers
     source          1. .211
                     /organism="Bos taurus"
                     /mol_type="genomic DNA"
                     /isolate="fetus 1"
                     /db_xref="taxon:9913"
                     /clone="1"
                     /tissue_type="brain"
                     /dev_stage="fetus"
                     /rearranged
     gene            1. .211
                     /gene="IGHV"
     CDS             <1. .>211
                     /gene="IGHV"
                     /codon_start=2
                     /product="immunoglobulin heavy chain variable region"
                     /protein_id="CAD55895.1"
                     /db_xref="GI:24459873"
                     /translation="PGLKSRLSITKDNSKSQVSLSVSSVTTEDTATYYCSYCTTVHQI
                     SCPDGYSYGYGCGYGYGCSGYDCYGY"
     exon            <1. .211
                     /gene="IGHV"
                     /number=1
     V_region        <1. .>211
                     /gene="IGHV"
                     /product="immunoglobulin heavy chain variable region"
     V_segment       1. .114
                     /gene="IGHV"
     J_segment       115. .211
                     /gene="IGHV"
ORIGIN

Alignment Scores:
Pred. No.:                0.0288          Length:         211
Score:                    78.00           Matches:        11
Percent Similarity:       100.00%         Conservative:   1
Best Local Similarity:    91.67%          Mismatches:     0
Query Match:              92.86%          Indels:         0
DB:                       4               Gaps:           0

US-09-858-332G-15 (1-12) x BTA512977 (1-211)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              :::||||||||||||||||||||||||||||||||||
Db        188 TACAACCATAACCATAACCACAACCATAACCATAAC 153

RESULT 144
AY095308/c
LOCUS       AY095308                244 bp    DNA     linear   PLN 09-SEP-2003
DEFINITION  Amaranthus palmeri x Amaranthus rudis clone 21D1.244f hybrid
            progeny specific sequence.
ACCESSION   AY095308
VERSION     AY095308.1  GI:20977610
KEYWORDS    .
SOURCE      Amaranthus palmeri x Amaranthus rudis
  ORGANISM  Amaranthus palmeri x Amaranthus rudis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            Caryophyllales; Amaranthaceae; Amaranthus.
REFERENCE   1  (bases 1 to 244)
  AUTHORS   Steinau,A.N., Skinner,D.Z. and Steinau,M.
  TITLE     Mechanism of Extreme Genetic Recombination in Weedy Amaranthus
            Hybrids
  JOURNAL   Weed Sci. 51 (5), 696-701 (2003)
REFERENCE   2  (bases 1 to 244)
  AUTHORS   Steinau,A.N. and Skinner,D.Z.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2002) Crop and Soil Sciences, Kansas State
            University, Washington State University, 209 Johnson Hall, Pullman,
            WA 99164-6420, USA
FEATURES             Location/Qualifiers
```

```
     source          1. .244
                     /organism="Amaranthus palmeri x Amaranthus rudis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:194672"
                     /clone="21D1.244f"
     misc_feature    1. .244
                     /note="unique to interspecific hybrid progeny; similar to
                     Arabidopsis thaliana mutator transposase gene encoded by
                     GenBank Accession Number AB009049"
ORIGIN

Alignment Scores:
Pred. No.:                0.0332          Length:          244
Score:                    78.00           Matches:         11
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              92.86%          Indels:          0
DB:                       15              Gaps:            0

US-09-858-332G-15 (1-12) x AY095308 (1-244)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              ||||||||||||||||||||||||||||||||||
Db        242 CACAATCACAATCACAATCACAATCACAATCAT 210

RESULT 145
AY095306/c
LOCUS       AY095306                263 bp    DNA     linear   PLN 09-SEP-2003
DEFINITION  Amaranthus palmeri x Amaranthus rudis clone 25d263f hybrid progeny
            specific sequence.
ACCESSION   AY095306
VERSION     AY095306.1  GI:20977608
KEYWORDS    .
SOURCE      Amaranthus palmeri x Amaranthus rudis
  ORGANISM  Amaranthus palmeri x Amaranthus rudis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            Caryophyllales; Amaranthaceae; Amaranthus.
REFERENCE   1  (bases 1 to 263)
  AUTHORS   Steinau,A.N., Skinner,D.Z. and Steinau,M.
  TITLE     Mechanism of Extreme Genetic Recombination in Weedy Amaranthus
            Hybrids
  JOURNAL   Weed Sci. 51 (5), 696-701 (2003)
REFERENCE   2  (bases 1 to 263)
  AUTHORS   Steinau,A.N. and Skinner,D.Z.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2002) Crop and Soil Sciences, Kansas State
            University, Washington State University, 209 Johnson Hall, Pullman,
            WA 99164-6420, USA
FEATURES             Location/Qualifiers
     source          1. .263
                     /organism="Amaranthus palmeri x Amaranthus rudis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:194672"
                     /clone="25d263f"
     misc_feature    1. .263
                     /note="unique to interspecific hybrid progeny; similar to
                     Arabidopsis thaliana mutator transposase gene encoded by
                     GenBank Accession Number AB009049"
ORIGIN

Alignment Scores:
Pred. No.:                0.0357          Length:          263
Score:                    78.00           Matches:         11
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              92.86%          Indels:          0
DB:                       15              Gaps:            0

US-09-858-332G-15 (1-12) x AY095306 (1-263)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              ||||||||||||||||||||||||||||||||||
Db        261 CACAATCACAATCACAATCACAATCACAATCAT 229

RESULT 146
AY095307/c
LOCUS       AY095307                267 bp    DNA     linear   PLN 09-SEP-2003
DEFINITION  Amaranthus palmeri x Amaranthus rudis clone 21a267f hybrid progeny
            specific sequence.
ACCESSION   AY095307
VERSION     AY095307.1  GI:20977609
KEYWORDS    .
SOURCE      Amaranthus palmeri x Amaranthus rudis
  ORGANISM  Amaranthus palmeri x Amaranthus rudis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            Caryophyllales; Amaranthaceae; Amaranthus.
REFERENCE   1  (bases 1 to 267)
  AUTHORS   Steinau,A.N., Skinner,D.Z. and Steinau,M.
  TITLE     Mechanism of Extreme Genetic Recombination in Weedy Amaranthus
            Hybrids
  JOURNAL   Weed Sci. 51 (5), 696-701 (2003)
REFERENCE   2  (bases 1 to 267)
  AUTHORS   Steinau,A.N. and Skinner,D.Z.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2002) Crop and Soil Sciences, Kansas State
            University, Washington State University, 209 Johnson Hall, Pullman,
            WA 99164-6420, USA
FEATURES             Location/Qualifiers
     source          1. .267
                     /organism="Amaranthus palmeri x Amaranthus rudis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:194672"
                     /clone="21a267f"
     misc_feature    1. .267
                     /note="unique to interspecific hybrid progeny; similar to
                     Arabidopsis thaliana mutator transposase gene encoded by
                     GenBank Accession Number AB009049"
ORIGIN

Alignment Scores:
Pred. No.:                0.0362          Length:          267
Score:                    78.00           Matches:         11
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              92.86%          Indels:          0
DB:                       15              Gaps:            0

US-09-858-332G-15 (1-12) x AY095307 (1-267)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              ||||||||||||||||||||||||||||||||||
Db        265 CACAATCACAATCACAATCACAATCACAATCAT 233

RESULT 147
AF097206/c
LOCUS       AF097206                381 bp    mRNA    linear   MAM 17-DEC-2003
DEFINITION  Bos taurus clone BF2D12 immunoglobulin heavy chain variable region
            mRNA, partial cds.
ACCESSION   AF097206
VERSION     AF097206.1  GI:3834652
KEYWORDS    .
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 381)
  AUTHORS   Saini,S.S. and Kaushik,A.
  TITLE     Extensive CDR3H length heterogeneity exists in bovine foetal VDJ
            rearrangements
  JOURNAL   Scand. J. Immunol. 55 (2), 140-148 (2002)
   PUBMED   11896930
REFERENCE   2  (bases 1 to 381)
  AUTHORS   Saini,S.S., Allore,B., Jacobs,R. and Kaushik,A.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1998) Pathobiology, University of Guelph, Guelph,
            ON N1G 2W1, Canada
FEATURES             Location/Qualifiers
     source          1. .381
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /clone="BF2D12"
                     /cell_type="mouse x Bovine heterohybridoma"
                     /dev_stage="fetal"
                     /note="PCR product"
     CDS             <1. .>381
                     /note="VDJ region"
                     /codon_start=1
                     /product="immunoglobulin heavy chain variable region"
                     /protein_id="AAC71041.1"
                     /db_xref="GI:3834653"
                     /translation="QVQLRESGPSLVKPSQTLSLTCTISGFSLSSYAVGWVRQAPGKA
                     LEWLGGISSGGSTCLNPALKSRLSITKDNSKSQVSLSVSSVTTEDTATYYCTCYSGYG
                     YGYGYAYGYDYVDAWGQGLLVTVSS"
ORIGIN

Alignment Scores:
Pred. No.:              0.051          Length:         381
Score:                  78.00          Matches:        11
Percent Similarity:     91.67%         Conservative:   0
Best Local Similarity:  91.67%         Mismatches:     1
Query Match:            92.86%         Indels:         0
DB:                     4              Gaps:           0

US-09-858-332G-15 (1-12) x AF097206 (1-381)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||||||||   |||||||||||||||||||||||||
Db        334 CATAACCATAAGCATAACCATAACCATAACCATAAC 299

RESULT 148
BTU49781/c
LOCUS       BTU49781                 408 bp    mRNA    linear   MAM 16-AUG-2001
DEFINITION  Bos taurus immunoglobulin rearranged heavy chain variable region
            mRNA, partial cds.
ACCESSION   U49781
VERSION     U49781.1  GI:1293607
KEYWORDS    .
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 408)
  AUTHORS   Berens,S.J., Wylie,D.E. and Lopez,O.J.
  TITLE     Use of a single VH family and long CDR3s in the variable region of
            cattle Ig heavy chains
  JOURNAL   Int. Immunol. 9 (1), 189-199 (1997)
   PUBMED   9043960
REFERENCE   2  (bases 1 to 408)
  AUTHORS   Berens,S.J., Wylie,D. and Lopez,O.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-1996) Osvaldo J. Lopez, Biological Sciences,
            University of Nebraska, 325 Manter Hall, Lincoln, NE 68588-0118,
            USA
FEATURES             Location/Qualifiers
     source          1. .408
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /cell_type="heterohybridoma"
                     /tissue_type="spleen"
                     /dev_stage="150 day fetus"
     CDS             <1. .>408
                     /codon_start=1
                     /product="immunoglobulin heavy chain variable region"
                     /protein_id="AAA98662.1"
                     /db_xref="GI:1293608"
                     /translation="FGLFTLSDKAVGWVRQAPGKALEWLGGIDTGGSTGYNPGLKSRL
                     SITKDNSKSQVSLSVSSVTTEDSATYYCTTVHQIRCPDGYGYGYGCGYGSYGYSGYDC
                     YGYGGYGGYGGYGGYSSYSYVDAWGQGLLVTVSS"
ORIGIN

Alignment Scores:
Pred. No.:              0.0544         Length:         408
Score:                  78.00          Matches:        11
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            92.86%         Indels:         0
DB:                     4              Gaps:           0

US-09-858-332G-15 (1-12) x BTU49781 (1-408)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 11
              |||||||||||||||||||||||||||||||||||
Db        277 CATAACCACAACCATAACCATAACCATAACCAT 245

RESULT 149
BTU50559/c
LOCUS       BTU50559                 480 bp    mRNA    linear   MAM 27-MAR-2002
DEFINITION  Bos taurus clone F1 IgM mRNA, partial cds.
ACCESSION   U50559
VERSION     U50559.1  GI:1399522
KEYWORDS    .
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 480)
  AUTHORS   Berens,S.J., Wylie,D. and Lopez,O.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-1996) Osvaldo J. Lopez, Biological Sciences,
            University of Nebraska, 325 Manter Hall, Lincoln, NE 68588-0118,
            USA
REFERENCE   2  (bases 1 to 480)
  AUTHORS   Berens,S.J., Wylie,D.E. and Lopez,O.J.
  TITLE     Use of a single VH family and long CDR3s in the variable region of
            cattle Ig heavy chains
  JOURNAL   Int. Immunol. 9 (1), 189-199 (1997)
   PUBMED   9043960
FEATURES             Location/Qualifiers
     source          1. .480
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /clone="F1"
                     /tissue_type="spleen"
                     /dev_stage="150 day fetus"
     CDS             <1. .>480
                     /note="rearranged VDJ exon; immunoglobulin"
                     /codon_start=1
                     /product="IgM"
                     /protein_id="AAB03276.1"
                     /db_xref="GI:1399523"
                     /translation="WRLVGVVTLICSKMNPLWTLLFVLSAPRGVLSQVQLRESGPSLV
                     KPSQTLSLTCTASGFSLSDKAVGWVRQAPGKALEWLGGIDTGGSTGYNPGLKSRLSIT
                     KDNSKSQVSLSVSSVTTEDSATYYCTTVHRSCYSGYGYGYGCGYGDAWGQGLLVTVSS
                     "
ORIGIN

Alignment Scores:
Pred. No.:              0.0637         Length:         480
Score:                  78.00          Matches:        11
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            92.86%         Indels:         0
DB:                     4              Gaps:           0
```

```
US-09-858-332G-15 (1-12) x BTU50559 (1-480)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              ||||||||||||||||||||||||||||||||||
Db        439 CATAACCACAACCATAACCATAACCATAACCAC 407

RESULT 150
AF097203/c
LOCUS       AF097203                489 bp    mRNA    linear   MAM 17-DEC-2003
DEFINITION  Bos taurus clone BF1H1 immunoglobulin heavy chain variable region
            mRNA, partial cds.
ACCESSION   AF097203
VERSION     AF097203.1  GI:3834646
KEYWORDS    .
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 489)
  AUTHORS   Saini,S.S. and Kaushik,A.
  TITLE     Extensive CDR3H length heterogeneity exists in bovine foetal VDJ
            rearrangements
  JOURNAL   Scand. J. Immunol. 55 (2), 140-148 (2002)
   PUBMED   11896930
REFERENCE   2  (bases 1 to 489)
  AUTHORS   Saini,S.S., Allore,B., Jacobs,R. and Kaushik,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1998) Pathobiology, University of Guelph, Guelph,
            ON N1G 2W1, Canada
FEATURES             Location/Qualifiers
     source          1. .489
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /clone="BF1H1"
                     /cell_type="mouse x Bovine heterohybridoma"
                     /dev_stage="fetal"
                     /note="PCR product"
     CDS             <1. .>489
                     /note="VDJ region"
                     /codon_start=1
                     /product="immunoglobulin heavy chain variable region"
                     /protein_id="AAC71038.1"
                     /db_xref="GI:3834647"
                     /translation="QVQLRESGPSLVKPSQTLSLTCTASGFSLSDKAVGWVRQAPGKA
                     LEWLGGIDTGGSTGYNPGLKSRLSITKDNSKSQVSLSVSSVTTEDSATYYCTTVHPSP
                     DGYSYGYGCGYGYGCSGYDCYGYGGYGYGGYGGYSSYSYSYSPRKYYVDAWGQGLLVT
                     VSS"
ORIGIN

Alignment Scores:
Pred. No.:              0.0648          Length:         489
Score:                  78.00           Matches:        11
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  91.67%          Mismatches:     0
Query Match:            92.86%          Indels:         0
DB:                     4               Gaps:           0

US-09-858-332G-15 (1-12) x AF097203 (1-489)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              :::||||||||||||||||||||||||||||||||||
Db        352 TACAACCATAACCATAACCACAACCATAACCATAAC 317

RESULT 151
AF097204/c
LOCUS       AF097204                492 bp    mRNA    linear   MAM 17-DEC-2003
DEFINITION  Bos taurus clone BF4E9 immunoglobulin heavy chain variable region
            mRNA, partial cds.
ACCESSION   AF097204
VERSION     AF097204.1  GI:3834648
KEYWORDS    .
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 492)
  AUTHORS   Saini,S.S. and Kaushik,A.
  TITLE     Extensive CDR3H length heterogeneity exists in bovine foetal VDJ
            rearrangements
  JOURNAL   Scand. J. Immunol. 55 (2), 140-148 (2002)
   PUBMED   11896930
REFERENCE   2  (bases 1 to 492)
  AUTHORS   Saini,S.S., Allore,B., Jacobs,R. and Kaushik,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1998) Pathobiology, University of Guelph, Guelph,
            ON N1G 2W1, Canada
FEATURES             Location/Qualifiers
     source          1. .492
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /clone="BF4E9"
                     /cell_type="mouse x Bovine heterohybridoma"
                     /dev_stage="fetal"
                     /note="PCR product"
     CDS             <1. .>492
                     /note="VDJ region"
                     /codon_start=1
                     /product="immunoglobulin heavy chain variable region"
                     /protein_id="AAC71039.1"
                     /db_xref="GI:3834649"
                     /translation="QVQLRESGPSLVKPSQTLSLTCTASGFSLSDKAVGWVRQAPGKA
                     LEWLGGIDTGGSTGYNPGLKSRLSITKDNSKSQVSLSVSSVTTEDSATYYCTTVHQIF
                     CPDGYSYGYGCGYGYGCSGYDCYGYGGYGYGGYGGYSSYSYSYSYEYYGDAWGQGLLV
                     TVSS"
ORIGIN

Alignment Scores:
Pred. No.:              0.0652          Length:         492
Score:                  78.00           Matches:        11
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  91.67%          Mismatches:     0
Query Match:            92.86%          Indels:         0
DB:                     4               Gaps:           0

US-09-858-332G-15 (1-12) x AF097204 (1-492)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              :::||||||||||||||||||||||||||||||||||
Db        358 TACAACCATAACCATAACCACAACCATAACCATAAC 323

RESULT 152
AR568827/c
LOCUS       AR568827                587 bp    DNA     linear   PAT 14-DEC-2004
DEFINITION  Sequence 1 from patent US 6740747.
ACCESSION   AR568827
VERSION     AR568827.1  GI:56568469
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 587)
  AUTHORS   Kaushik,A. and Saini,S.S.
  TITLE     Bovine VDJ cassette, BF1H1, suitable for antigenization
  JOURNAL   Patent: US 6740747-A 1 25-MAY-2004;
FEATURES             Location/Qualifiers
     source          1. .587
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              0.0773          Length:         587
```

```
Score:                 78.00         Matches:       11
Percent Similarity:    100.00%       Conservative:  1
Best Local Similarity: 91.67%        Mismatches:    0
Query Match:           92.86%        Indels:        0
DB:                    6             Gaps:          0

US-09-858-332G-15 (1-12) x AR568827 (1-587)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              :::||||||||||||||||||||||||||||||||||
Db        417 TACAACCATAACCATAACCACAACCATAACCATAAC 382

RESULT 153
AR546135
LOCUS       AR546135                726 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 1266 from patent US 6747137.
ACCESSION   AR546135
VERSION     AR546135.1  GI:53939310
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 726)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 1266 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .726
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:             0.0948        Length:        726
Score:                 78.00         Matches:       11
Percent Similarity:    91.67%        Conservative:  0
Best Local Similarity: 91.67%        Mismatches:    1
Query Match:           92.86%        Indels:        0
DB:                    6             Gaps:          0

US-09-858-332G-15 (1-12) x AR546135 (1-726)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||   ||||||||||||||||||
Db         46 CACAACCACAATCATGGTCATAACCATAATCACAAT 81

RESULT 154
AR449702/c
LOCUS       AR449702               1731 bp    DNA     linear   PAT 20-FEB-2004
DEFINITION  Sequence 366 from patent US 6673910.
ACCESSION   AR449702
VERSION     AR449702.1  GI:42678759
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1731)
  AUTHORS   Breton,G.L.
  TITLE     Nucleic acid and amino acid sequences relating to M. catarrhalis
            for diagnostics and therapeutics
  JOURNAL   Patent: US 6673910-A 366 06-JAN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .1731
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:             0.219         Length:        1731
Score:                 78.00         Matches:       11
Percent Similarity:    100.00%       Conservative:  0
Best Local Similarity: 100.00%       Mismatches:    0
Query Match:           92.86%        Indels:        0
DB:                    6             Gaps:          0

US-09-858-332G-15 (1-12) x AR449702 (1-1731)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              ||||||||||||||||||||||||||||||||
Db       1344 CACAATCATAATCATAATCATAATCACAACCAT 1312

RESULT 155
AR527958
LOCUS       AR527958               2580 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 261 from patent US 6723837.
ACCESSION   AR527958
VERSION     AR527958.1  GI:53915961
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2580)
  AUTHORS   Karunanandaa,B., Yu,J. and Kishore,G.
  TITLE     Nucleic acid molecule and encoded protein associated with sterol
            synthesis and metabolism
  JOURNAL   Patent: US 6723837-A 261 20-APR-2004;
            Monsanto Technology LLC; St. Louis, MO
FEATURES             Location/Qualifiers
     source          1. .2580
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:             0.322         Length:        2580
Score:                 78.00         Matches:       11
Percent Similarity:    100.00%       Conservative:  0
Best Local Similarity: 100.00%       Mismatches:    0
Query Match:           92.86%        Indels:        0
DB:                    6             Gaps:          0

US-09-858-332G-15 (1-12) x AR527958 (1-2580)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              ||||||||||||||||||||||||||||||||
Db       1900 CATAATCATAATCATAATCATAATCATAACCAT 1932

RESULT 156
AR146444
LOCUS       AR146444               3455 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 3 from patent US 6218512.
ACCESSION   AR146444
VERSION     AR146444.1  GI:15109633
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3455)
  AUTHORS   Guarente,L.P., Austriaco,N. Jr. and Kennedy,B.
  TITLE     Genes determining cellular senescence in yeast
  JOURNAL   Patent: US 6218512-A 3 17-APR-2001;
FEATURES             Location/Qualifiers
     source          1. .3455
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

Alignment Scores:
Pred. No.:             0.426         Length:        3455
Score:                 78.00         Matches:       11
Percent Similarity:    100.00%       Conservative:  0
Best Local Similarity: 100.00%       Mismatches:    0
Query Match:           92.86%        Indels:        0
```

```
DB:                    6              Gaps:          0

US-09-858-332G-15 (1-12) x AR146444 (1-3455)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              ||||||||||||||||||||||||||||||||||
Db       2487 CATAATCATAATCATAATCATAATCATAACCAT 2519

RESULT 157
AR579950
LOCUS       AR579950                3455 bp    DNA     linear   PAT 15-DEC-2004
DEFINITION  Sequence 3 from patent US 6787300.
ACCESSION   AR579950
VERSION     AR579950.1  GI:56609846
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3455)
  AUTHORS   Guarente,L.P., Austriaco,N. Jr. and Kennedy,B.
  TITLE     Identifying lifespan-altering agents
  JOURNAL   Patent: US 6787300-A 3 07-SEP-2004;
            Massachusetts Institute of Technology; Cambridge, MA
FEATURES             Location/Qualifiers
     source          1. .3455
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:                0.426          Length:          3455
Score:                    78.00          Matches:         11
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              92.86%         Indels:          0
DB:                       6              Gaps:            0

US-09-858-332G-15 (1-12) x AR579950 (1-3455)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              ||||||||||||||||||||||||||||||||||
Db       2487 CATAATCATAATCATAATCATAATCATAACCAT 2519

RESULT 158
YSCMPT5
LOCUS       YSCMPT5                 3455 bp    DNA     linear   PLN 06-FEB-1999
DEFINITION  Yeast MPT5 gene for suppressor protein, complete cds.
ACCESSION   D26184
VERSION     D26184.1  GI:450489
KEYWORDS    multicopy suppressor; suppressor protein.
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 3455)
  AUTHORS   Sakai,A.
  TITLE     Multicopy suppressors of the yeast pop2 mutation
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3455)
  AUTHORS   Sakai,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-DEC-1993) Akira Sakai, Mitsubishi Kasei Institute of
            Life Sciences, Molecular Biology; 11, Minamiooya, Machida, Tokyo
            194, Japan (Tel:0427-24-6258, Fax:0427-29-1252)
COMMENT     Submitted (23-Dec-1993) to DDBJ by:
            Akira Sakai
            Department of Molecular Genetics
            Mitsubishi Kasei Institute of Life Sicneces
            11 Minamiooya, Machida-shi
            Tokyo 194
            Japan
            Phone:  0427-24-6258
            Fax:    0427-24-6316.

FEATURES             Location/Qualifiers
     source          1. .3455
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /strain="S288C"
                     /db_xref="taxon:4932"
     gene            663. .3167
                     /gene="MPT5"
     CDS             663. .3167
                     /gene="MPT5"
                     /note="multicopy suppressor of pop2 mutation"
                     /codon_start=1
                     /product="suppressor protein"
                     /protein_id="BAA05172.1"
                     /db_xref="GI:450490"
                     /translation="MSYNHQPQLSINSVQSLLEPVTPPPLGQMNNKRNHQKAHSLDLS
                     GFNQFISSTQSPLALMNNTSTSNSANSFSPNPNAASNSTGLSASMANPPAILPLINEF
                     DLEMDGPRRKSSHDFTVVAPSNSGVNTSSLIMETPSSSVTPAASLRNFSNSNNAASKC
                     GVDNSSFGLSSSTSSSMVEISALPLRDLDYIKLATDQFGCRFLQKKLETPSESNMVRD
                     LMYEQIKPFFLDLILDPFGNYLVQKLCDYLTAEQKTLLIQTIYPNVFQISINQYGTRS
                     LQKIIDTVDNEVQIDLIIKGFSQEFTSIEQVVTLINDLNGNHVIQKCIFKFSPSKFGF
                     IIDAIVEQNNIITISTHKHGCCVLQKLLSVCTLQQIFKISVKIVQFLPGLINDQFGNY
                     IIQFLLDIKELDFYLLAELFNRLSNELCQLSCLKFSSNVVEKFIKKLFRIITGFIVNN
                     NGGASQRTAVASDDVINASMNILLTTIDIFTVNLNVLIRDNFGNYALQTLLDVKNYSP
                     LLAYNKNSNAIGQNSSSTLNYGNFCNDFSLKIGNLIVLTKELLPSIKTTSYAKKIKLK
                     VKAYAEATGIPFTDISPQVTAMSHNNLQTINNENKNPHNKNSHNHNHNHNHNHAHNNN
                     NNNNQKSHTRHFSLPANAYHRRSNSSVTNNFSNQYAQDQKIHSPQQIMNFNQNAYPSM
                     GAPSFNSQTNPPLVSHNSLQNFDNRQFANLMAHPNSAAPIHSFSSSNITNVNPNVSRG
                     FKQPGFMMNETDKINANHFSPYSNANSQNFNESFVPRMQYQTEGANWDSSLSMKSQHI
                     GQGPYNQVNMSRNASISNMPAMNTARTSDELQFTLP"
ORIGIN

Alignment Scores:
Pred. No.:                0.426          Length:          3455
Score:                    78.00          Matches:         11
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              92.86%         Indels:          0
DB:                       15             Gaps:            0

US-09-858-332G-15 (1-12) x YSCMPT5 (1-3455)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              ||||||||||||||||||||||||||||||||||
Db       2487 CATAATCATAATCATAATCATAATCATAACCAT 2519

RESULT 159
YSCHTR1
LOCUS       YSCHTR1                 3769 bp    DNA     linear   PLN 29-MAY-2002
DEFINITION  Saccharomyces cerevisiae HTR1 gene, complete cds.
ACCESSION   D25541
VERSION     D25541.1  GI:436256
KEYWORDS    .
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 3769)
  AUTHORS   Kikuchi,Y., Oka,Y., Kobayashi,M., Uesono,Y., Toh-e,A. and
            Kikuchi,A.
  TITLE     A new yeast gene, HTR1, required for growth at high temperature, is
            needed for recovery from mating pheromone-induced G1 arrest
  JOURNAL   Mol. Gen. Genet. 245 (1), 107-116 (1994)
   PUBMED   7845352
REFERENCE   2  (bases 1 to 3769)
  AUTHORS   Kikuchi,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-NOV-1993) Akihiko Kikuchi, Mitsubishi Kasei Institute
            of Life Sciences; 11 Minamiooya, Machida, Tokyo 194, Japan
            (Tel:0427-24-6229, Fax:0427-24-6317)
COMMENT     Submitted (26-Nov-1993) to DDBJ by:
            Akihiko Kikuchi
            Mitsubishi Kasei Institute of Life Sciences
```

```
          11 Minamiooya, Machida-shi
          Tokyo 194
          Japan
          Phone: 0427-24-6229
          Fax:   0427-24-6317.
FEATURES             Location/Qualifiers
     source          1. .3769
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /strain="ATCC70391"
                     /db_xref="taxon:4932"
                     /chromosome="12"
                     /map="12 L"
     gene            665. .3169
                     /gene="HTR1"
     CDS             665. .3169
                     /gene="HTR1"
                     /codon_start=1
                     /protein_id="BAA05024.1"
                     /db_xref="GI:469512"
                     /translation="MSYNHQPQLSINSVQSLLEPVTPPPLGQMNNKRNHQKAHSLDLS
                     GFNQFISSTQSPLALMNNTSTSNSANSFSPNPNAASNSTGLSASMANPPAILPLINEF
                     DLEMDGPRRKSSHDFTVVAPSNSGVNTSSLIMETPSSSVTPAASLRNFSNSNNAASKC
                     GVDNSSFGLSSSTSSSMVEISALPLRDLDYIKLATDQFGCRFLQKKLETPSESNMVRD
                     LMYEQIKPFFLDLILDPFGNYLVQKLCDYLTAEQKTLLIQTIYPNVFQISINQYGTRS
                     LQKIIDTVDNEVQIDLIIKGFSQEFTSIEQVVTLINDLNGNHVIQKCIFKFSPSKFGF
                     IIDAIVEQNNIITISTHKHGCCVLQKLLSVCTLQQIFKISVKIVQFLPGLINDQFGNY
                     IIQFLLDIKELDFYLLAELFNRLSNELCQLSCLKFSSNVVEKFIKKLFRIITGFIVNN
                     NGGASQRTAVASDDVINASMNILLTTIDIFTVNLNVLIRDNFGNYALQTLLDVKNYSP
                     LLAYNKNSNAIGQNSSSTLNYGNFCNDFSLKIGNLIVLTKELLPSIKTTSYAKKIKLK
                     VKAYAEATGIPFTDISPQVTAMSHNNLQTINNENKNPHNKNSHNHNHNHNHNHAHNNN
                     NNNNQKSHTRHFSLPANAYHRRSNSSVTNNFSNQYAQDQKIHSPQQIMNFNQNAYPSM
                     GAPSFNSQTNPPLVSHNSLQNFDNRQFANLMAHPNSAAPIHSFSSSNITNVNPNVSRG
                     FKQPGFMMNETDKINANHFSPYSNANSQNFNESFVPRMQYQTEGANWDSSLSMKSQHI
                     GQGPYNQVNMSRNASISNMPAMNTARTSDELQFTLP"
ORIGIN

Alignment Scores:
Pred. No.:                0.463          Length:          3769
Score:                    78.00          Matches:         11
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              92.86%         Indels:          0
DB:                       15             Gaps:            0

US-09-858-332G-15 (1-12) x YSCHTR1 (1-3769)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              ||||||||||||||||||||||||||||||||||
Db       2489 CATAATCATAATCATAATCATAATCATAACCAT 2521

RESULT 160
SCYGL176C
LOCUS       SCYGL176C               5378 bp    DNA     linear   PLN 11-AUG-1997
DEFINITION  S.cerevisiae chromosome VII reading frame ORF YGL176c.
ACCESSION   Z72698 Y13135
VERSION     Z72698.1  GI:1322783
KEYWORDS    .
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 5378)
  AUTHORS   Bruschi,C.V., Coglievina,M., Bertani,I., Klima,R., Zaccaria,P. and
            Delneri,D.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 5378)
  AUTHORS   MIPS.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAY-1996) Data collected by MIPS on behalf of the
            European yeast chromosome VII sequencing project. MIPS at the
            Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
            Martinsried, FRG; E-mail: Mewes@mips.embnet.org
FEATURES             Location/Qualifiers
     source          1. .5378
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4932"
                     /chromosome="VII"
     gene            1. .2665
                     /gene="MPT5"
     intron          <1. .88
                     /gene="MPT5"
                     /note="ORF YGL178w"
                     /number=1
     exon            89. .2665
                     /gene="MPT5"
                     /note="ORF YGL178w"
                     /number=2
     CDS             complement(3509. .5173)
                     /note="unnamed protein product; ORF YGL176c"
                     /codon_start=1
                     /protein_id="CAA96888.1"
                     /db_xref="GI:1322784"
                     /db_xref="SGD:S0003144"
                     /db_xref="UniProt/Swiss-Prot:P46945"
                     /translation="MTPNLPGFYYDRERRRYFRISENQSISTTGTTNQYRKDNIKRQC
                     VEENYDKKFSMIKKKRQQTLQKYKLSLLNPLERAFRPLSYEKYMIGLNMQYASHSLTE
                     GHHSHSSANVKSLNFPHRIQIGVLANCILLVTQEGCFHSKLVFATNKGYVAGFSSLDN
                     FSEENFFTGFSMAELNPMLKYKSEPTDVFKTMKLERTVAIKEGPSHYFYHNVNTRSNV
                     HTFAIFLQDFSSLKLLKIRQVKLKENCQVHDSLVVGDTLIITVNYRCHFYDLIPETFP
                     NPYIFSPAKSSRKHKSRSDITSLSFCLQEDALSPLKKTNTGVFYLGYRNGDSMAIVFT
                     NITNMTLQYSKTNGMTSESRNQPIRNSLKSVVSIKALNNKGLILISGMADKENVQQLV
                     IADTFLEDILTEIPVVSFKTKFLNVTKDTEILEISDDGRYFIYGSTSARDGKGDFEVF
                     CTTLSGNLDYEKSEGGNITLYPIGGMKNYCRLENFQFESIHLHSAFIPPRYVNPFDAV
                     EPLGEESSTSPYDIPEEALSQKICILIRREDDPYNGANIFITSALT"
ORIGIN

Alignment Scores:
Pred. No.:                0.652          Length:          5378
Score:                    78.00          Matches:         11
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              92.86%         Indels:          0
DB:                       15             Gaps:            0

US-09-858-332G-15 (1-12) x SCYGL176C (1-5378)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              ||||||||||||||||||||||||||||||||||
Db       1985 CATAATCATAATCATAATCATAATCATAACCAT 2017

RESULT 161
SCYGL178W
LOCUS       SCYGL178W               6323 bp    DNA     linear   PLN 18-APR-2005
DEFINITION  S.cerevisiae chromosome VII reading frame ORF YGL178w.
ACCESSION   Z72700 Y13135
VERSION     Z72700.1  GI:1322785
KEYWORDS    .
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 6323)
  AUTHORS   Bruschi,C.V., Coglievina,M., Bertani,I., Klima,R., Zaccaria,P. and
            Delneri,D.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 6323)
  AUTHORS   MIPS.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAY-1996) Data collected by MIPS on behalf of the
            European yeast chromosome VII sequencing project. MIPS at the
            Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
            Martinsried, FRG; E-mail: Mewes@mips.embnet.org
FEATURES             Location/Qualifiers
     source          1. .6323
```

```
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4932"
                     /chromosome="VII"
     gene            join(2261. .2263,2904. .5480)
                     /gene="MPT5"
     CDS             join(2261. .2263,2904. .5480)
                     /gene="MPT5"
                     /note="ORF YGL178w"
                     /codon_start=1
                     /protein_id="CAA96889.1"
                     /db_xref="GI:1322786"
                     /db_xref="GOA:P39016"
                     /db_xref="SGD:S000003146"
                     /db_xref="UniProt/Swiss-Prot:P39016"
                     /translation="MINNEPFPSADSASILTTSTSNNSLMSYNHQPQLSINSVQSLLE
                     PVTPPPLGQMNNKRNHQKAHSLDLSGFNQFISSTQSPLALMNNTSTSNSANSFSPNPN
                     AASNSTGLSASMANPPAILPLINEFDLEMDGPRRKSSHDFTVVAPSNSGVNTSSLIME
                     TPSSSVTPAASLRNFSNSNNAASKCGVDNSSFGLSSSTSSSMVEISALPLRDLDYIKL
                     ATDQFGCRFLQKKLETPSESNMVRDLMYEQIKPFFLDLILDPFGNYLVQKLCDYLTAE
                     QKTLLIQTIYPNVFQISINQYGTRSLQKIIDTVDNEVQIDLIIKGFSQEFTSIEQVVT
                     LINDLNGNHVIQKCIFKFSPSKFGFIIDAIVEQNNIITISTHKHGCCVLQKLLSVCTL
                     QQIFKISVKIVQFLPGLINDQFGNYIIQFLLDIKELDFYLLAELFNRLSNELCQLSCL
                     KFSSNVVEKFIKKLFRIITGFIVNNNGGASQRTAVASDDVINASMNILLTTIDIFTVN
                     LNVLIRDNFGNYALQTLLDVKNYSPLLAYNKNSNAIGQNSSSTLNYGNFCNDFSLKIG
                     NLIVLTKELLPSIKTTSYAKKIKLKVKAYAEATGIPFTDISPQVTAMSHNNLQTINNE
                     NKNPHNKNSHNHNHNHNHNHAHNNNNNNNQKSHTRHFSLPANAYHRRSNSSVTNNFSN
                     QYAQDQKIHSPQQIMNFNQNAYPSMGAPSFNSQTNPPLVSHNSLQNFDNRQFANLMAH
                     PNSAAPIHSFSSSNITNVNPNVSRGFKQPGFMMNETDKINANHFSPYSNANSQNFNES
                     FVPRMQYQTEGANWDSSLSMKSQHIGQGPYNQVNMSRNASISNMPAMNTARTSDELQF
                     TLP"
     exon            <2261. .2263
                     /gene="MPT5"
                     /number=1
     intron          2264. .2903
                     /gene="MPT5"
                     /number=1
     CDS             2468. .2815
                     /gene="MPT5"
                     /note="unnamed protein product; ORF YGL177w"
                     /codon_start=1
                     /protein_id="CAA96890.1"
                     /db_xref="GI:1322787"
                     /db_xref="SGD:S000003145"
                     /db_xref="UniProt/Swiss-Prot:P53105"
                     /translation="MPIRIFSKEGLRERGASGKNEQKKKKKEKIRRSMKNLCRFVESY
                     KDFTLYGYRFHSKIKHKFCVLSSSFPVLFFCHPLSPLKSNFIDRFTLDSHASQATWTL
                     YMVVIVLFLSYRN"
     exon            2904. .>5480
                     /gene="MPT5"
                     /number=2
ORIGIN

Alignment Scores:
Pred. No.:              0.763          Length:         6323
Score:                  78.00          Matches:        11
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            92.86%         Indels:         0
DB:                     15             Gaps:           0

US-09-858-332G-15 (1-12) x SCYGL178W (1-6323)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              ||||||||||||||||||||||||||||||||
Db       4800 CATAATCATAATCATAATCATAATCATAACCAT 4832

RESULT 162
SCDNAGMH
LOCUS       SCDNAGMH                7951 bp    DNA     linear   PLN 18-APR-2005
DEFINITION  S.cerevisiae G1618 gene and MPT5/HTR1 gene.
ACCESSION   X83690
VERSION     X83690.1  GI:794143
KEYWORDS    G1618 gene; HTR1 gene; MPT5 gene; mutation suppressor; ser/thr
            protein kinase; upstream activation sequence.
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1
  AUTHORS   Coglievina,M., Bertani,I., Klima,R., Zaccaria,P. and Bruschi,C.V.
  TITLE     The DNA sequence of a 7941 bp fragment of the left arm of
            chromosome VII of Saccharomyces cerevisiae contains four open
            reading frames including the multicopy suppressor gene of the pop2
            mutation and a putative serine/threonine protein kinase gene
  JOURNAL   Yeast 11 (8), 767-774 (1995)
   PUBMED   7668046
REFERENCE   2  (bases 1 to 7951)
  AUTHORS   Bruschi,C.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-1995) C.V. Bruschi, Microbiology, International
            Centre for Genetic Engineering and Biotechnology, AREA Science
            Park, Padriciano 99, I-34012 Trieste, ITALY
COMMENT     Sequence overlapping with those under the accession numbers D26184
            and D25541 and also with est 102637 of the TIGR database.
FEATURES             Location/Qualifiers
     source          1. .7951
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /strain="FY1679"
                     /db_xref="taxon:4932"
                     /chromosome="VII"
                     /map="left arm"
     gene            complement(87. .1769)
                     /gene="biE560"
     CDS             complement(87. .1769)
                     /gene="biE560"
                     /codon_start=1
                     /product="putative ser /thr protein kinases"
                     /protein_id="CAA58659.1"
                     /db_xref="GI:794144"
                     /db_xref="GOA:P43637"
                     /db_xref="InterPro:IPR000719"
                     /db_xref="InterPro:IPR002290"
                     /db_xref="InterPro:IPR008271"
                     /db_xref="InterPro:IPR011009"
                     /db_xref="SGD:S000003147"
                     /db_xref="UniProt/Swiss-Prot:P43637"
                     /translation="MVLLKEPVQPLPRSSLLYNNASNSSSRIKETRKVKLLYNPLTKR
                     QILNNFEILATLGNGQYGKVKLARDLGTGALVAIKILNRFEKRSGYSLQLKVENPRVN
                     QEIEVMKRCHHENVVELYEILNDPESTKVYLVLEYCSRGPVKWCPENKMEIKAVGPSI
                     LTFQQSRKVVLDVVSGLEYLHSQGITHRDIKPSNLLISSNGTVKISDFGVAMSTATGS
                     TNIQSSHEQLLKSRALGTPAFFAPELCSTEKEYSCSSAIDIWSLGVTIYCLLFGKLPF
                     NANSGLELFDSIINKPLEFPSYEEMLNGATSGITMEEYTDAKDLLKKLLQKDPDKRIK
                     LADIKVHPFMCHYGKSDAASVLTNLETFHELKVSPPSSCKRVELVSLPVNSSFASLDS
                     VYMENFDHNNLRTGADRNSTYSPSIYDANTLSPSAYHNIGSRESSYSSFSSFTSSTAF
                     ASQISIQDAPAIGDQQCLIGESGSSLRVNSCEFPQYTTMSPVGEYPFESTEASLSSTL
                     TPVGNVPQRIKAHLVEGKSNSKDDLRIEADASLVFEASDAQRTRRRMSLYKL"
     gene            1000. .1377
                     /gene="biB125"
     CDS             1000. .1377
                     /gene="biB125"
                     /note="putative protein"
                     /codon_start=1
                     /protein_id="CAA58661.1"
                     /db_xref="GI:1209469"
                     /db_xref="UniProt/TrEMBL:Q05671"
                     /translation="MADEQEYSFSVEQSSGAKKAGVPKALDFKSCSCDDCILVLPVAV
                     DIATPKSDIFTVPFDEINRFDGLISRCVIPWECKYSRPDTTSNTTFLDCWKVSIDGPT
                     AFISILFSGHHLTGPLEQYSNTR"
     gene            1881. .1888
                     /gene="MPT5=HTR1"
     enhancer        1881. .1888
                     /gene="MPT5=HTR1"
                     /note="HAP2/3/4 putative UAS"
     gene            join(4030. .4032,4673. .7073,7912. .7946)
```

```
                /gene="MPT5=HTR1"
     CDS        join(4030. .4032,4673. .7073,7912. .7946)
                /gene="MPT5=HTR1"
                /function="multicopy suppressor of pop2 mutation"
                /codon_start=1
                /protein_id="CAA58660.1"
                /db_xref="GI:794145"
                /db_xref="GOA:P39016"
                /db_xref="UniProt/Swiss-Prot:P39016"
                /translation="MINNEPPPSADSASILTTSTSNNSLMSYNHQPQLSINSVQSLLE
                PVTPPPLGQMNNKRNHQKAHSLDLSGFNQFISSTQSPLALMNNTSTSNSANSFSPNPN
                AASNSTGLSASMANPPAILPLINEFDLEMDGPRRKSSHDFTVVAPSNSGVNTSSLIME
                TPSSSVTPAASLRNFSNSNNAASKCGVDNSSFGLSSSTSSSMVEISALPLRDLDYIKL
                ATDQFGCRFLQKKLETPSESNMVRDLMYEQIKPFFLDLILDPFGNYLVQKLCDYLTAE
                QKTLLIQTIYPNVFQISINQYGTRSLQKIIDTVDNEVQIDLIIKGFSQEFTSIEQVVT
                LINDLNGNHVIQKCIFKFSPSKFGFIIDAIVEQNNIITISTHKHGCCVLQKLLSVCTL
                QQIFKISVKIVQFLPGLINDQFGNYIIQFLLDIKELDFYLLAELFNRLSNELCQLSCL
                KFSSNVVEKFIKKLFRIITGFIVNNNGGASQRTAVASDDVINASMNILLTTIDIFTVN
                LNVLIRDNFGNYALQTLLDVKNYSPLLAYNKNSNAIGQNSSSTLNYGNFCNDFSLKIG
                NLIVLTKELLPSIKTTSYAKKIKLKVKAYAEATGIPFTDISPQVTAMSHNNLQTINNE
                NKNPHNKNSHNHNHNHNHNHAHNNNNNNNQKSHTRHFSLPANAYHRRSNSSVTNNFSN
                QYAQDQKIHSPQQIMNFNQNAYPSMGAPSFNSQTNPPLVSHNSLQNFDNRQFANLMAH
                PNSAAPIHSFSSSNITNVNPNVSRGFKQPGFMMNETDKINANHFSPYSNANSQNFNES
                FVPQWNCTPRRIML"
     exon       4030. .4032
                /gene="MPT5=HTR1"
                /number=1
     intron     4033. .4672
                /gene="MPT5=HTR1"
                /number=1
     gene       4237. .4584
                /gene="biB115"
     CDS        4237. .4584
                /gene="biB115"
                /note="putative protein"
                /codon_start=1
                /protein_id="CAA58662.1"
                /db_xref="GI:1209470"
                /db_xref="SGD:S000003145"
                /db_xref="UniProt/Swiss-Prot:P53105"
                /translation="MPIRIFSKEGLRERGASGKNEQKKKKKEKIRRSMKNLCRFVESY
                KDFTLYGYRFHSKIKHKFCVLSSSFPVLFFCHPLSPLKSNFIDRFTLDSHASQATWTL
                YMVVIVLFLSYRN"
     misc_feature 4643. .4650
                /gene="MPT5=HTR1"
                /note="splicing branchpoint"
     exon       4673. .7073
                /gene="MPT5=HTR1"
                /number=2
     gene       4745. .7249
                /gene="bic834"
     CDS        4745. .7249
                /gene="bic834"
                /note="putative protein"
                /codon_start=1
                /protein_id="CAA58663.1"
                /db_xref="GI:1209471"
                /db_xref="GOA:P39016"
                /db_xref="InterPro:IPR001313"
                /db_xref="UniProt/Swiss-Prot:P39016"
                /translation="MSYNHQPQLSINSVQSLLEPVTPPPLGQMNNKRNHQKAHSLDLS
                GFNQFISSTQSPLALMNNTSTSNSANSFSPNPNAASNSTGLSASMANPPAILPLINEF
                DLEMDGPRRKSSHDFTVVAPSNSGVNTSSLIMETPSSSVTPAASLRNFSNSNNAASKC
                GVDNSSFGLSSSTSSSMVEISALPLRDLDYIKLATDQFGCRFLQKKLETPSESNMVRD
                LMYEQIKPFFLDLILDPFGNYLVQKLCDYLTAEQKTLLIQTIYPNVFQISINQYGTRS
                LQKIIDTVDNEVQIDLIIKGFSQEFTSIEQVVTLINDLNGNHVIQKCIFKFSPSKFGF
                IIDAIVEQNNIITISTHKHGCCVLQKLLSVCTLQQIFKISVKIVQFLPGLINDQFGNY
                IIQFLLDIKELDFYLLAELFNRLSNELCQLSCLKFSSNVVEKFIKKLFRIITGFIVNN
                NGGASQRTAVASDDVINASMNILLTTIDIFTVNLNVLIRDNFGNYALQTLLDVKNYSP
                LLAYNKNSNAIGQNSSSTLNYGNFCNDFSLKIGNLIVLTKELLPSIKTTSYAKKIKLK
                VKAYAEATGIPFTDISPQVTAMSHNNLQTINNENKNPHNKNSHNHNHNHNHNHAHNNN
                NNNNQKSHTRHFSLPANAYHRRSNSSVTNNFSNQYAQDQKIHSPQQIMNFNQNAYPSM
                GAPSFNSQTNPPLVSHNSLQNFDNRQFANLMAHPNSAAPIHSFSSSNITNVNPNVSRG
                FKQPGFMMNETDKINANHFSPYSNANSQNFNESFVPRMQYQTEGANWDSSLSMKSQHI
                GQGPYNQVNMSRNASISNMPAMNTARTSDELQFTLP"
     intron     7074. .7911
                /gene="MPT5=HTR1"
                /number=2
     misc_feature 7835. .7842
                /gene="MPT5=HTR1"
                /note="splicing branchpoint"
     exon       7912. .7946
                /gene="MPT5=HTR1"
                /number=3
ORIGIN

Alignment Scores:
Pred. No.:              0.951          Length:          7951
Score:                  78.00          Matches:         11
Percent Similarity:     100.00%        Conservative:    0
Best Local Similarity:  100.00%        Mismatches:      0
Query Match:            92.86%         Indels:          0
DB:                     15             Gaps:            0

US-09-858-332G-15 (1-12) x SCDNAGMH (1-7951)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              ||||||||||||||||||||||||||||||||
Db       6569 CATAATCATAATCATAATCATAATCATAACCAT 6601

RESULT 163
AC126793/c
LOCUS       AC126793              109787 bp    DNA     linear   PLN 24-JUN-2004
DEFINITION  Medicago truncatula clone mth2-17m16, complete sequence.
ACCESSION   AC126793
VERSION     AC126793.13  GI:49170206
KEYWORDS    HTG.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 109787)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Medicago truncatula BAC Clone mth2-17m16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 109787)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 109787)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-SEP-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   4  (bases 1 to 109787)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   5  (bases 1 to 109787)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JAN-2004) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
```

```
REFERENCE   6  (bases 1 to 109787)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUN-2004) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Jun 24, 2004 this sequence version replaced gi:41393273.
            -------------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ----------------------.
FEATURES             Location/Qualifiers
     source          1. .109787
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3880"
                     /clone="mth2-17m16"
                     /clone_lib="Medicago truncatula BAC library H2"
ORIGIN

Alignment Scores:
Pred. No.:                11.9          Length:         109787
Score:                    78.00         Matches:        11
Percent Similarity:       91.67%        Conservative:   0
Best Local Similarity:    91.67%        Mismatches:     1
Query Match:              92.86%        Indels:         0
DB:                       15            Gaps:           0

US-09-858-332G-15 (1-12) x AC126793 (1-109787)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||   ||||||||||||||||||||||||||||||
Db      55319 CATACTCATAATCATAATCATAATCATAATCATAAT 55284

RESULT 164
CR382128_04
WPCOMMENT
  Sequence split into 31 fragments  LOCUS CR382128 Accession CR382128
     Fragment Name      Begin     End
     CR382128_00            1   110000
     CR382128_01       100001   210000
     CR382128_02       200001   310000
     CR382128_03       300001   410000
     CR382128_04       400001   510000
     CR382128_05       500001   610000
     CR382128_06       600001   710000
     CR382128_07       700001   810000
     CR382128_08       800001   910000
     CR382128_09       900001  1010000
     CR382128_10      1000001  1110000
     CR382128_11      1100001  1210000
     CR382128_12      1200001  1310000
     CR382128_13      1300001  1410000
     CR382128_14      1400001  1510000
     CR382128_15      1500001  1610000
     CR382128_16      1600001  1710000
     CR382128_17      1700001  1810000
     CR382128_18      1800001  1910000
     CR382128_19      1900001  2010000
     CR382128_20      2000001  2110000
     CR382128_21      2100001  2210000
     CR382128_22      2200001  2310000
     CR382128_23      2300001  2410000
     CR382128_24      2400001  2510000
     CR382128_25      2500001  2610000
     CR382128_26      2600001  2710000
     CR382128_27      2700001  2810000
     CR382128_28      2800001  2910000
     CR382128_29      2900001  3010000
     CR382128_30      3000001  3066374
Continuation (5 of 31) of CR382128 from base 400001 (CR382128 Yarrowia lipolytica chromo

Alignment Scores:
Pred. No.:                11.9          Length:         110000
Score:                    78.00         Matches:        11
Percent Similarity:       91.67%        Conservative:   0
Best Local Similarity:    91.67%        Mismatches:     1
Query Match:              92.86%        Indels:         0
DB:                       15            Gaps:           0

US-09-858-332G-15 (1-12) x CR382128_04 (1-110000)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||   ||||||||||||||||||||||||
Db      65842 CACAACCACACCCACAATCACAACCACAACCACAAC 65877

RESULT 165
AP008209_138/c
WPCOMMENT
  Sequence split into 362 fragments  LOCUS AP008209 Accession AP008209
     Fragment Name      Begin     End
     AP008209_000           1   110000
     AP008209_001      100001   210000
     AP008209_002      200001   310000
     AP008209_003      300001   410000
     AP008209_004      400001   510000
     AP008209_005      500001   610000
     AP008209_006      600001   710000
     AP008209_007      700001   810000
     AP008209_008      800001   910000
     AP008209_009      900001  1010000
     AP008209_010     1000001  1110000
     AP008209_011     1100001  1210000
     AP008209_012     1200001  1310000
     AP008209_013     1300001  1410000
     AP008209_014     1400001  1510000
     AP008209_015     1500001  1610000
     AP008209_016     1600001  1710000
     AP008209_017     1700001  1810000
     AP008209_018     1800001  1910000
     AP008209_019     1900001  2010000
     AP008209_020     2000001  2110000
     AP008209_021     2100001  2210000
     AP008209_022     2200001  2310000
     AP008209_023     2300001  2410000
     AP008209_024     2400001  2510000
     AP008209_025     2500001  2610000
     AP008209_026     2600001  2710000
     AP008209_027     2700001  2810000
     AP008209_028     2800001  2910000
     AP008209_029     2900001  3010000
     AP008209_030     3000001  3110000
     AP008209_031     3100001  3210000
     AP008209_032     3200001  3310000
     AP008209_033     3300001  3410000
     AP008209_034     3400001  3510000
     AP008209_035     3500001  3610000
     AP008209_036     3600001  3710000
     AP008209_037     3700001  3810000
     AP008209_038     3800001  3910000
     AP008209_039     3900001  4010000
     AP008209_040     4000001  4110000
     AP008209_041     4100001  4210000
     AP008209_042     4200001  4310000
     AP008209_043     4300001  4410000
     AP008209_044     4400001  4510000
     AP008209_045     4500001  4610000
     AP008209_046     4600001  4710000
     AP008209_047     4700001  4810000
     AP008209_048     4800001  4910000
     AP008209_049     4900001  5010000
     AP008209_050     5000001  5110000
     AP008209_051     5100001  5210000
     AP008209_052     5200001  5310000
```

AP008209_053	5300001	5410000
AP008209_054	5400001	5510000
AP008209_055	5500001	5610000
AP008209_056	5600001	5710000
AP008209_057	5700001	5810000
AP008209_058	5800001	5910000
AP008209_059	5900001	6010000
AP008209_060	6000001	6110000
AP008209_061	6100001	6210000
AP008209_062	6200001	6310000
AP008209_063	6300001	6410000
AP008209_064	6400001	6510000
AP008209_065	6500001	6610000
AP008209_066	6600001	6710000
AP008209_067	6700001	6810000
AP008209_068	6800001	6910000
AP008209_069	6900001	7010000
AP008209_070	7000001	7110000
AP008209_071	7100001	7210000
AP008209_072	7200001	7310000
AP008209_073	7300001	7410000
AP008209_074	7400001	7510000
AP008209_075	7500001	7610000
AP008209_076	7600001	7710000
AP008209_077	7700001	7810000
AP008209_078	7800001	7910000
AP008209_079	7900001	8010000
AP008209_080	8000001	8110000
AP008209_081	8100001	8210000
AP008209_082	8200001	8310000
AP008209_083	8300001	8410000
AP008209_084	8400001	8510000
AP008209_085	8500001	8610000
AP008209_086	8600001	8710000
AP008209_087	8700001	8810000
AP008209_088	8800001	8910000
AP008209_089	8900001	9010000
AP008209_090	9000001	9110000
AP008209_091	9100001	9210000
AP008209_092	9200001	9310000
AP008209_093	9300001	9410000
AP008209_094	9400001	9510000
AP008209_095	9500001	9610000
AP008209_096	9600001	9710000
AP008209_097	9700001	9810000
AP008209_098	9800001	9910000
AP008209_099	9900001	10010000
AP008209_100	10000001	10110000
AP008209_101	10100001	10210000
AP008209_102	10200001	10310000
AP008209_103	10300001	10410000
AP008209_104	10400001	10510000
AP008209_105	10500001	10610000
AP008209_106	10600001	10710000
AP008209_107	10700001	10810000
AP008209_108	10800001	10910000
AP008209_109	10900001	11010000
AP008209_110	11000001	11110000
AP008209_111	11100001	11210000
AP008209_112	11200001	11310000
AP008209_113	11300001	11410000
AP008209_114	11400001	11510000
AP008209_115	11500001	11610000
AP008209_116	11600001	11710000
AP008209_117	11700001	11810000
AP008209_118	11800001	11910000
AP008209_119	11900001	12010000
AP008209_120	12000001	12110000
AP008209_121	12100001	12210000
AP008209_122	12200001	12310000
AP008209_123	12300001	12410000
AP008209_124	12400001	12510000
AP008209_125	12500001	12610000
AP008209_126	12600001	12710000
AP008209_127	12700001	12810000
AP008209_128	12800001	12910000
AP008209_129	12900001	13010000
AP008209_130	13000001	13110000
AP008209_131	13100001	13210000
AP008209_132	13200001	13310000
AP008209_133	13300001	13410000
AP008209_134	13400001	13510000
AP008209_135	13500001	13610000
AP008209_136	13600001	13710000
AP008209_137	13700001	13810000
AP008209_138	13800001	13910000
AP008209_139	13900001	14010000
AP008209_140	14000001	14110000
AP008209_141	14100001	14210000
AP008209_142	14200001	14310000
AP008209_143	14300001	14410000
AP008209_144	14400001	14510000
AP008209_145	14500001	14610000
AP008209_146	14600001	14710000
AP008209_147	14700001	14810000
AP008209_148	14800001	14910000
AP008209_149	14900001	15010000
AP008209_150	15000001	15110000
AP008209_151	15100001	15210000
AP008209_152	15200001	15310000
AP008209_153	15300001	15410000
AP008209_154	15400001	15510000
AP008209_155	15500001	15610000
AP008209_156	15600001	15710000
AP008209_157	15700001	15810000
AP008209_158	15800001	15910000
AP008209_159	15900001	16010000
AP008209_160	16000001	16110000
AP008209_161	16100001	16210000
AP008209_162	16200001	16310000
AP008209_163	16300001	16410000
AP008209_164	16400001	16510000
AP008209_165	16500001	16610000
AP008209_166	16600001	16710000
AP008209_167	16700001	16810000
AP008209_168	16800001	16910000
AP008209_169	16900001	17010000
AP008209_170	17000001	17110000
AP008209_171	17100001	17210000
AP008209_172	17200001	17310000
AP008209_173	17300001	17410000
AP008209_174	17400001	17510000
AP008209_175	17500001	17610000
AP008209_176	17600001	17710000
AP008209_177	17700001	17810000
AP008209_178	17800001	17910000
AP008209_179	17900001	18010000
AP008209_180	18000001	18110000
AP008209_181	18100001	18210000
AP008209_182	18200001	18310000
AP008209_183	18300001	18410000
AP008209_184	18400001	18510000
AP008209_185	18500001	18610000
AP008209_186	18600001	18710000
AP008209_187	18700001	18810000
AP008209_188	18800001	18910000
AP008209_189	18900001	19010000
AP008209_190	19000001	19110000
AP008209_191	19100001	19210000
AP008209_192	19200001	19310000
AP008209_193	19300001	19410000
AP008209_194	19400001	19510000
AP008209_195	19500001	19610000
AP008209_196	19600001	19710000

Alignment Scores:

Pred. No.:	11.9	Length:	110000
Score:	78.00	Matches:	11
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	91.67%	Mismatches:	0
Query Match:	92.86%	Indels:	0
DB:	15	Gaps:	0

US-09-858-332G-15 (1-12) x AP008209_138 (1-110000)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              :::||||||||||||||||||||||||||||||||
Db     105595 TATAATCATAATCATAATCATAATCATAATCATAAT 105560
```

RESULT 166
AP008209_139/c
WPCOMMENT
Sequence split into 362 fragments LOCUS AP008209 Accession AP008209

Fragment Name	Begin	End
AP008209_000	1	110000
AP008209_001	100001	210000
AP008209_002	200001	310000
AP008209_003	300001	410000
AP008209_004	400001	510000
AP008209_005	500001	610000
AP008209_006	600001	710000
AP008209_007	700001	810000
AP008209_008	800001	910000
AP008209_009	900001	1010000
AP008209_010	1000001	1110000
AP008209_011	1100001	1210000
AP008209_012	1200001	1310000
AP008209_013	1300001	1410000
AP008209_014	1400001	1510000
AP008209_015	1500001	1610000
AP008209_016	1600001	1710000
AP008209_017	1700001	1810000
AP008209_018	1800001	1910000
AP008209_019	1900001	2010000
AP008209_020	2000001	2110000
AP008209_021	2100001	2210000
AP008209_022	2200001	2310000
AP008209_023	2300001	2410000
AP008209_024	2400001	2510000
AP008209_025	2500001	2610000
AP008209_026	2600001	2710000
AP008209_027	2700001	2810000
AP008209_028	2800001	2910000
AP008209_029	2900001	3010000
AP008209_030	3000001	3110000
AP008209_031	3100001	3210000
AP008209_032	3200001	3310000
AP008209_033	3300001	3410000
AP008209_034	3400001	3510000
AP008209_035	3500001	3610000
AP008209_036	3600001	3710000
AP008209_037	3700001	3810000
AP008209_038	3800001	3910000
AP008209_039	3900001	4010000
AP008209_040	4000001	4110000
AP008209_041	4100001	4210000
AP008209_042	4200001	4310000
AP008209_043	4300001	4410000
AP008209_044	4400001	4510000
AP008209_045	4500001	4610000
AP008209_046	4600001	4710000
AP008209_047	4700001	4810000
AP008209_048	4800001	4910000
AP008209_049	4900001	5010000
AP008209_050	5000001	5110000
AP008209_051	5100001	5210000
AP008209_052	5200001	5310000
AP008209_053	5300001	5410000
AP008209_054	5400001	5510000
AP008209_055	5500001	5610000
AP008209_056	5600001	5710000
AP008209_057	5700001	5810000
AP008209_058	5800001	5910000
AP008209_059	5900001	6010000
AP008209_060	6000001	6110000
AP008209_061	6100001	6210000
AP008209_062	6200001	6310000
AP008209_063	6300001	6410000
AP008209_064	6400001	6510000
AP008209_065	6500001	6610000
AP008209_066	6600001	6710000
AP008209_067	6700001	6810000
AP008209_068	6800001	6910000
AP008209_069	6900001	7010000
AP008209_070	7000001	7110000
AP008209_071	7100001	7210000
AP008209_072	7200001	7310000
AP008209_073	7300001	7410000
AP008209_074	7400001	7510000
AP008209_075	7500001	7610000
AP008209_076	7600001	7710000
AP008209_077	7700001	7810000
AP008209_078	7800001	7910000
AP008209_079	7900001	8010000
AP008209_080	8000001	8110000
AP008209_081	8100001	8210000
AP008209_082	8200001	8310000
AP008209_083	8300001	8410000
AP008209_084	8400001	8510000
AP008209_085	8500001	8610000
AP008209_086	8600001	8710000
AP008209_087	8700001	8810000
AP008209_088	8800001	8910000
AP008209_089	8900001	9010000
AP008209_090	9000001	9110000
AP008209_091	9100001	9210000
AP008209_092	9200001	9310000
AP008209_093	9300001	9410000
AP008209_094	9400001	9510000
AP008209_095	9500001	9610000
AP008209_096	9600001	9710000
AP008209_097	9700001	9810000
AP008209_098	9800001	9910000
AP008209_099	9900001	10010000
AP008209_100	10000001	10110000
AP008209_101	10100001	10210000
AP008209_102	10200001	10310000
AP008209_103	10300001	10410000
AP008209_104	10400001	10510000
AP008209_105	10500001	10610000
AP008209_106	10600001	10710000
AP008209_107	10700001	10810000
AP008209_108	10800001	10910000
AP008209_109	10900001	11010000
AP008209_110	11000001	11110000
AP008209_111	11100001	11210000
AP008209_112	11200001	11310000
AP008209_113	11300001	11410000
AP008209_114	11400001	11510000
AP008209_115	11500001	11610000
AP008209_116	11600001	11710000
AP008209_117	11700001	11810000
AP008209_118	11800001	11910000
AP008209_119	11900001	12010000
AP008209_120	12000001	12110000
AP008209_121	12100001	12210000
AP008209_122	12200001	12310000
AP008209_123	12300001	12410000
AP008209_124	12400001	12510000
AP008209_125	12500001	12610000
AP008209_126	12600001	12710000
AP008209_127	12700001	12810000

AP008209_128	12800001	12910000
AP008209_129	12900001	13010000
AP008209_130	13000001	13110000
AP008209_131	13100001	13210000
AP008209_132	13200001	13310000
AP008209_133	13300001	13410000
AP008209_134	13400001	13510000
AP008209_135	13500001	13610000
AP008209_136	13600001	13710000
AP008209_137	13700001	13810000
AP008209_138	13800001	13910000
AP008209_139	13900001	14010000
AP008209_140	14000001	14110000
AP008209_141	14100001	14210000
AP008209_142	14200001	14310000
AP008209_143	14300001	14410000
AP008209_144	14400001	14510000
AP008209_145	14500001	14610000
AP008209_146	14600001	14710000
AP008209_147	14700001	14810000
AP008209_148	14800001	14910000
AP008209_149	14900001	15010000
AP008209_150	15000001	15110000
AP008209_151	15100001	15210000
AP008209_152	15200001	15310000
AP008209_153	15300001	15410000
AP008209_154	15400001	15510000
AP008209_155	15500001	15610000
AP008209_156	15600001	15710000
AP008209_157	15700001	15810000
AP008209_158	15800001	15910000
AP008209_159	15900001	16010000
AP008209_160	16000001	16110000
AP008209_161	16100001	16210000
AP008209_162	16200001	16310000
AP008209_163	16300001	16410000
AP008209_164	16400001	16510000
AP008209_165	16500001	16610000
AP008209_166	16600001	16710000
AP008209_167	16700001	16810000
AP008209_168	16800001	16910000
AP008209_169	16900001	17010000
AP008209_170	17000001	17110000
AP008209_171	17100001	17210000
AP008209_172	17200001	17310000
AP008209_173	17300001	17410000
AP008209_174	17400001	17510000
AP008209_175	17500001	17610000
AP008209_176	17600001	17710000
AP008209_177	17700001	17810000
AP008209_178	17800001	17910000
AP008209_179	17900001	18010000
AP008209_180	18000001	18110000
AP008209_181	18100001	18210000
AP008209_182	18200001	18310000
AP008209_183	18300001	18410000
AP008209_184	18400001	18510000
AP008209_185	18500001	18610000
AP008209_186	18600001	18710000
AP008209_187	18700001	18810000
AP008209_188	18800001	18910000
AP008209_189	18900001	19010000
AP008209_190	19000001	19110000
AP008209_191	19100001	19210000
AP008209_192	19200001	19310000
AP008209_193	19300001	19410000
AP008209_194	19400001	19510000
AP008209_195	19500001	19610000
AP008209_196	19600001	19710000

Alignment Scores:

Pred. No.:	11.9	Length:	110000
Score:	78.00	Matches:	11
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	91.67%	Mismatches:	0
Query Match:	92.86%	Indels:	0
DB:	15	Gaps:	0

US-09-858-332G-15 (1-12) x AP008209_139 (1-110000)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              :::||||||||||||||||||||||||||||||||
Db       5595 TATAATCATAATCATAATCATAATCATAATCATAAT 5560
```

RESULT 167
NCB13H18

```
LOCUS       NCB13H18              122576 bp    DNA     linear   PLN 16-APR-2005
DEFINITION  Neurospora crassa DNA linkage group V BAC contig B13H18.
ACCESSION   AL807374
VERSION     AL807374.1  GI:21622501
KEYWORDS    .
SOURCE      Neurospora crassa
  ORGANISM  Neurospora crassa
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
REFERENCE   1
  AUTHORS   Schulte,U., Aign,V., Hoheisel,J., Brandt,P., Fartmann,B.,
            Holland,R., Nyakatura,G., Mewes,H.W. and Mannhaupt,G.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 122576)
  AUTHORS   German Neurospora genome,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2002) MIPS, Institut fuer Bioinformatik,
            GSF-Forschungszentrum fuer Umwelt und Gesundheit, GmbH,
            Ingolstaedter Landstrasse 1, D-85764 Neuherberg, FRG, E-mail:
            G.Mannhaupt@gsf.de Project Coordinator: Ulrich Schulte, Institute
            of Biochemistry, Heinrich-Heine-University, D-40225 Duesseldorf,
            E-mail: ulrich-schulte@uni-duesseldorf.de
COMMENT     this contig is an assembly of BAC clones 13H18 from 1 to: 65723,
            9O3 from 65724 to: 92584 (strain OR74A) and cosmid 108G5 from 92584
            to: 122576 rev, cosmid library pLORIST6Xh, strain 74-OR-23-1A; BAC
            and cosmid clones are available at the Fungal Genetic Stock Center,
            http://www.fgsc.net
            Sequencing was performed by MWG Biotech AG, Ebersberg, Germany,
            http://www.mwgdna.com
            Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of linkage groups II
            and V can be viewed at: http://mips.gsf.de/proj/neurospora.
FEATURES             Location/Qualifiers
     source          1. .122576
                     /organism="Neurospora crassa"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5141"
                     /chromosome="2"
     misc_feature    1. .6
                     /note="overlap to Cosmid contig 80A10"
     gene            1122. .4574
                     /gene="B13H18.010"
     gene            complement(1122. .4574)
                     /gene="B13H18.010"
     CDS             complement(1122. .4574)
                     /gene="B13H18.010"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAD37049.1"
                     /db_xref="GI:21622502"
                     /db_xref="UniProt/TrEMBL:Q8NIV8"
                     /translation="MSLGVKQLWRVLSNPNGARGLSLIPRECTPALARAAPTSSRSTL
                     RALHNQAPSLPDQSQPDDAEFEANDVRRRQHTTTEIHDSGENPAPAAQTPTEPSRASP
                     PPPPPPLPREEAQEQAHTQAGDVQDCQNDSSPAKKAKKAKPAKPKRGRGSGLVNRLAK
                     VPVRAHPVLESVPVRPVTEYPDQYETRGSVRTHMKDNHRQKMVMGTRVPSNRTDWRAV
                     LRNLLQSTPMYLPEQFADNVKVIIPRETAPRLLSDSEYSVWDIKSRTSCGMVLYRAGD
                     AGYEEGTEDQQPYLIIWGHHHSVNTAIDEILGVTRKVTIISRTKGTEKVLWNGKVDSK
                     EPIVFVSAHRTPSQLAPYHVNIRADEFPKPEEWTRKTFEEYVDALIKSRMPSALASKL
                     YPEHGWHEMAVIEQLVAVFNDEVAAQFVTNAAFKMALHYMTRGGETWRPEVLSMFNYA
                     KSGRVRLDVGVFNILAEAAVKTRSLYRFSAVLYLLVTHGYQPNLRTWILFLRMFEAEE
```

```
                VKRYVLQAMHDRGLFNARGALRMVADEMAPHDAYRAVQLGWDVPTFLAKQDELYGANR
                KWVSIDALNKIFHVFGSYSMFAEIRQLLDLIFAGKLIAYPDHVTVNTILTHCRAQKKL
                DLAVEFIQLFEGHRAKNNETPLKLDNLACDILFDLAQWHKKPHVLSTIWRYAHLVNST
                TFHMRQKGTILLKMDAGKLKEEKPLKRLKLGDEERVLFVRNLLLGEFIQANGGDEVFG
                KIIAGIAKREQKHQQKMEEIDDDDRIPPSLGDWFHSQSLLDPSSSPTISPSSSSAAAS
                LPASSSNPDHNRSSILQPQSQSSANEPESTTWGNIYGSFRSWSFASKYLHLEPSAPIG
                SLLQQALDRDRDLHLALRTEGLDRHAIEELMKPVEIPTKLRPKPAFEEVKKIFKREGE
                VTKRGVKKVGVKEGDRDGDEEVQEAMALGPVDGQALDNWLNGGGSGNEVVPAEKNAHV
                DETAATKPQPTSLPDESKKVVVETQKQEQQQQQLTAATVEAEVEAEVEVELEAKGSPE
                NKVKEEEEAHKRAEKETTTTTSKSKTSPAASASASASAPAPAPAPTQPTTSPLTTS
                SPLPTPPTTSPQSTKPTTKPPTKERKTKEEAANLKLRFPSSSSSSSSSSSGKRVKWED
                LADL"
exon            complement(1122. .4574)
                /gene="B13H18.010"
                /number=1
repeat_region   1344. .1373
                /note="30 bp tggtgctggtgctggtgc tandem repeat"
repeat_region   1371. .1397
                /note="27 bp tgctgatgctgatgc tandem repeat"
repeat_region   4250. .4273
                /note="24 bp gtggtggtggtg tandem repeat"
repeat_region   5317. .5348
                /note="32 bp tcactcactcactcactcac tandem repeat"
gene            5368. .8490
                /gene="B13H18.020"
gene            complement(join(5368. .8215,8289. .8365,8455. .8490))
                /gene="B13H18.020"
CDS             complement(join(5368. .8215,8289. .8365,8455. .8490))
                /gene="B13H18.020"
                /note="similarity to OSBP-related protein 7 mRNA, Homo
                sapiens, TREMBL:AF323729_1
                contains EST gb:BG278417, BG278418, AI392604, AI398171"
                /codon_start=1
                /product="related to OSBP-related protein 7"
                /protein_id="CAD37050.1"
                /db_xref="GI:21622503"
                /db_xref="GOA:Q8NIV7"
                /db_xref="InterPro:IPR000348"
                /db_xref="InterPro:IPR000648"
                /db_xref="InterPro:IPR001849"
                /db_xref="UniProt/TrEMBL:Q8NIV7"
                /translation="MAGIEQLEIHSKSYIVRWVKVDEGHTISWSLQPHKKSINFAIVK
                HPGSGGTAAGSQTFEATATPEHPATEAAVSDAKSGLFRRDASTAQDQLAKKGFIPIQW
                HGKCEADKVTVGTYDVTSAGMYGLVFDNTFSKQTSKTATFVLLTYPTGAPPQTAHHLP
                TVQAGAGSNASRTSLGAGTTPRLGAANSDSVDSLPSNETAGRGRAASNARSEGGFVPG
                ATYHVGVLLKRRRKKGQGFARRFFSLDFTTCTLSYYHNRNSSALRGAIPLSLAAIAAD
                ERRREITIDSGAEIWHLKASGPKDFNDWAKALEKASRIARGLETLPNGASLGVDRPAS
                RPSTSVPSAQQEDEREWAQVESLVSRVAGTRDALRRLVKDMAAEKQAPGASQASLLSP
                NTPAIAEDSETYFTPSTESRRPFWKRKTSNASPGAQPGVQATAGAALAVPSPASATAV
                AEERNPYDNCAALLNDLDSVVVEFSQLLSKNKRRRIPSQMMPPNMMPAASRKSYESTA
                STIDEFFDAEAGEQGGAGQNQVMIIRQSEEDMAGSDAEEVDIHDSSSVSSVEEEEEFE
                GVGTENHDSLFPVKPKSLAPLPITDTVTRRTTIPAAKVPPPSLIAFVRKNVGKDLSTI
                SMPVSANEPLSLCQRMAEQLEYGQLLTEAAKKTDANERLLYVTAFAVSQFSNGRAKER
                ATRKPFNPLLGETFELVRSENEVPGGFRVLVEKICHRPVRLAVQADSALWSFSQSPAP
                SQKFWGKSAEITTEGRVRVSLHLPDGTDELYSWHIATVFLRNVVYGEKYVEPVGNMNV
                NNETSGAKAVVEFRSAKGMFGGRGEEVHVETFGPDGSNTGQALTGNWTSSLKTAPGGK
                EIWKAGSLVPNAANTYGFTTFAASLNEITPLEKDKLPPTDSRLRPDQRFHEQGDLDSA
                EEWKVKLEEAQRVRRKQMEERGEEYKPKWFVKVATGQDGEEVWKLKGGKEGYWEERAK
                SGGQWTGVLDIFSG"
exon            complement(5368. .8215)
                /gene="B13H18.020"
                /number=1
intron          complement(8216. .8288)
                /gene="B13H18.020"
                /number=1
exon            complement(8289. .8365)
                /gene="B13H18.020"
                /number=2
intron          complement(8366. .8454)
                /gene="B13H18.020"
                /number=2
exon            complement(8455. .8490)
                /gene="B13H18.020"
                /number=3
gene            9972. .14390
                /gene="B13H18.030"
gene            complement(join(9972. .10231,10299. .13896,13967. .14149,
                14223. .14390))
                /gene="B13H18.030"
CDS             complement(join(9972. .10231,10299. .13896,13967. .14149,
                14223. .14390))
                /gene="B13H18.030"
                /note="similarity to SPT6 protein, YGR116W, Saccharomyces
                cerevisiae, PIR:A36468
                contains EST gb:AA897934"
                /codon_start=1
                /product="related to transcriptional regulator protein
                SPT6"
                /protein_id="CAD37051.1"
                /db_xref="GI:21622504"
                /db_xref="GOA:Q8NIV6"
                /db_xref="InterPro:IPR000980"
                /db_xref="InterPro:IPR003029"
                /db_xref="UniProt/TrEMBL:Q8NIV6"
                /translation="MSNSMRDLIDGEAELDDEEDDESFDEEAGDRPRRRPNIDDSSEE
                EEDDEDEEEARKIREGFIVDEDEEDEAEDSDARERRRRKKRRREREEEEQLDEEDLDL
                IGEAIPEWERKPQPQRLKRGHRDDHRPTERRGLAEIFSDEDEEHDDRGYGRPSGRAQA
                DEFDDFIEDDYPEDDEERRHREEDEEVARPKDRGLNIDTTGLDKDALEDMDAIFGNGE
                DYEWALQLEEEQEHAERTKEDIELQDVFEPSQLKEKLLTDEDNRIRFNDEPERFQLDR
                KAFKNLQMTSDQFKEEARWISNLMLPSKNLSSELHGPFNKAVGKVLEFFVIDGVEVPY
                VFQHRRDYLIHAKKMRNPNRRDDPDAPEYTVDAEKLLTQDDLWKVLDLDIRFRSFLEK
                RNALEQTYDKLKEKTRDDILEEMIRQAQSIEELQDLQDYLNFQYSAELKDLAANDNSA
                QREIKRAGGRTAQFERIRRSNAYKFVQALGITPDRLAKNILRESSKVTSEDDSRLPDD
                LADTLVDADFPTGELVINAARQMLAEEMFASPRMRKHFRKNFYGMGIVSCRRTDKGLR
                KIDEANPYYEVKYLKNMSIADLAVRPELFLKMMKAEEEGLIEIKVSLENDREFRQQLF
                SDFASENFSELADKWNAERQKVIDLAFDKLVKVIVKGVKDSLRTACQDELLKTCRELY
                FKRLDQAPYKPKGMVIGTTPRVLTLSNGMGDPNREPVSWVSMDEDGRILEHGTFTNLA
                RDESQREALAELVRRRQPDVIGISGFSADTHRLIKDVEGLVSEKGLVGPEYDDPETNE
                YRSDLLEVIVINDEVARLYKDSPRAVADHPSLNPMTRYCIALARYMQNPMKEYAALGK
                DVTSLQIHPYQQYLPQAKLLKHLETAMVDMVNLVGVDINVAMQDANTAHLLPYVAGLG
                PRKAQLLIKGINKNGGVVTSRDELVGDPERHKLPVLGPRVWNNCASFLFIEYEPTNPE
                SDPLDNTRIHPEDYDLARKVAADALGLDEEDVKAETDENGAGAIVRKLFKDDEQDKVN
                ELILEEYAEQLEREYQQRKRATLETIRAELQVPYEELRKKFESLTVDQVFTMLTGENR
                DSLCEGMIVAANVRVVKDDFAIVKLDCGIEGRIESHDVSYRHSIKDVLHVGQVVQAKL
                IDLNRKEFVSKLSMRDEEMRRPFRRHFDHGRDQWDYRKEDEDREELREKDKSTGRAQR
                VVNHPLFKPFNSTQAEEYLGSQPSGEVVIRPSSKGNDHLAVTWKVADGVFQHVDVLEL
                QKENEFAVGRVLRVGKYTYQDLDELIVDHVKAMAKKVDELMQCDKFQKGSRNETEKWL
                TTYMDANPNRSTYAFCIDTKHPGYFFLCFKASRNSRVNAWPVRVIPHAFELMKSQYPD
                VRALCNGFKLRYQSEMLKQQSGGR"
exon            complement(9972. .10231)
                /gene="B13H18.030"
                /number=1
intron          complement(10232. .10298)
                /gene="B13H18.030"
                /number=1
exon            complement(10299. .13896)
                /gene="B13H18.030"
                /number=2
intron          complement(13897. .13966)
                /gene="B13H18.030"
                /number=2
exon            complement(13967. .14149)
                /gene="B13H18.030"
                /number=3
intron          complement(14150. .14222)
                /gene="B13H18.030"
                /number=3
exon            complement(14223. .14390)
                /gene="B13H18.030"
                /number=4

Alignment Scores:
Pred. No.:              13.3            Length:         122576
Score:                  78.00           Matches:        11
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
```

```
Query Match:            92.86%      Indels:        0
DB:                     15          Gaps:          0

US-09-858-332G-15 (1-12) x NCB13H18 (1-122576)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              ||||||||||||||||||||||||||||||||||
Db      57190 CACAACCACAACCACAACCACAACCACAACCAC 57222

RESULT 168
AC135907
LOCUS       AC135907               161676 bp    DNA     linear   PLN 24-OCT-2002
DEFINITION  Genomic sequence for Oryza sativa, Nipponbare strain, clone
            OJ1057D12, from chromosome 3, complete sequence.
ACCESSION   AC135907
VERSION     AC135907.1  GI:24308636
KEYWORDS    HTG.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 161676)
  AUTHORS   McCombie,W.R., de la Bastide,M., Spiegel,L., Preston,R.,
            Nascimento,L., Zutavern,T., Balija,V., Bell,M., Miller,B.,
            Katzenberger,F., Muller,S., Sullivan,P., Yang,C., Dike,S.,
            O'Shaughnessy,A., Palmer,L. and Dedhia,N.
  TITLE     Genomic sequence for Oryza sativa, Nipponbare strain, clone
            OJ1057D12, from chromosome 3, complete sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 161676)
  AUTHORS   McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT-2002) Lita Annenberg Hazen Genome Center, Cold
            Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor,
            NY 11724, USA
COMMENT     This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest.  The nucleotide
            sequence of this BAC clone was generated by combining Monsanto,
            Syngenta and Cold Spring Harbor Laboratory Genome Center sequencing
            data.
            Clone OJ1057D12 overlaps clone OSJNBa0048D11 (AC107619) from base
            129099 to base 161676.  The overlap is from base 126362 to base
            158911 (complemented) on OSJNBa0048D11.
FEATURES             Location/Qualifiers
     source          1. .161676
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /strain="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="3"
                     /clone="OJ1057D12"
                     /clone_lib="HindIII"
                     /note="(japonica cultivar-group)"
     misc_feature    12969. .13074
                     /note="The sequence is covered by a single high quality
                     plasmid subclone which was sequenced with dye terminator
                     chemistry.  There is coverage by two additional subclones
                     in the region with quality below phred30."
     misc_feature    143125. .143214
                     /note="The sequence is a dinucleotide repeat (TA) in which
                     the exact number of repeat copies is unknown.  Sequences
                     in OJ1057D12 show up to 36 fewer TA pairs than the
                     sequences in the overlapping clone OSJNBa0048D11
                     (AC107619).  The expanded version is shown in the
                     assembly."
     misc_feature    143403. .143458
                     /note="The sequence is a dinucleotide (TA) repeat in which
                     the exact number of TA pairs is unknown.  Sequences in
                     OJ1057D12 show up to 14 additional TA pairs compared to
                     sequences in the overlapping clone OSJNBa0048D11
                     (AC107619).  The expanded version is shown in the
                     assembly."
     misc_feature    144029. .144174
                     /note="The sequence is covered solely by an overlap with
                     clone OSJNBa0048D11 (AC107619)."
     misc_feature    157631. .157834
                     /note="The sequence is covered solely by an overlap with
                     clone OSJNBa0048D11 (AC107619)."
ORIGIN

Alignment Scores:
Pred. No.:              17.3          Length:        161676
Score:                  78.00         Matches:       11
Percent Similarity:     100.00%       Conservative:  1
Best Local Similarity:  91.67%        Mismatches:    0
Query Match:            92.86%        Indels:        0
DB:                     15            Gaps:          0

US-09-858-332G-15 (1-12) x AC135907 (1-161676)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              :::||||||||||||||||||||||||||||||||||
Db     100862 TATAATCATAATCATAATCATAATCATAATCATAAT 100897

RESULT 169
AC116420
LOCUS       AC116420               191497 bp    DNA     linear   ROD 04-FEB-2004
DEFINITION  Mus musculus chromosome 1, clone RP24-141B16, complete sequence.
ACCESSION   AC116420
VERSION     AC116420.8  GI:42269807
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 191497)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 1, clone RP24-141B16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 191497)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 191497)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
```

```
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JAN-2004) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 191497)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-FEB-2004) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 4, 2004 this sequence version replaced gi:40737081.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L25089
                Center clone name: 141_B_16
            --------------.
FEATURES             Location/Qualifiers
     source          1. .191497
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="1"
                     /map="1"
                     /clone="RP24-141B16"
                     /clone_lib="RPCI-24 Male Mouse BAC"
     misc_feature    1. .8250
                     /note="wgs_end_extension
                     clone_end:SP6"
     repeat_region   59. .90
                     /rpt_family="(A)n"
     repeat_region   complement(351. .420)
                     /rpt_family="MTE"
     repeat_region   complement(435. .873)
                     /rpt_family="MER69B"
     repeat_region   complement(1611. .1734)
                     /rpt_family="RMER1A"
     repeat_region   1735. .1760
                     /rpt_family="(CCA)n"
     repeat_region   complement(1761. .3373)
                     /rpt_family="RMER1A"
     repeat_region   3955. .4010
                     /rpt_family="(CA)n"
     repeat_region   complement(4165. .4584)
                     /rpt_family="Lx"
     repeat_region   complement(5590. .5733)
                     /rpt_family="Charlie2"
     repeat_region   complement(7787. .8120)
                     /rpt_family="ORR1D"
     misc_feature    8251. .8254
                     /note="clone_boundary
                     clone_end:SP6
                     site:MboI"
     repeat_region   complement(9681. .10275)
                     /rpt_family="Lx8"
     repeat_region   10926. .11002
                     /rpt_family="MLT1L"
     repeat_region   11130. .11280
                     /rpt_family="CT-rich"
     repeat_region   11471. .11508
                     /rpt_family="(TG)n"
     repeat_region   12223. .12602
                     /rpt_family="L1MA5"
     repeat_region   12730. .12931
                     /rpt_family="L1MA5"
     repeat_region   12949. .13079
                     /rpt_family="RSINE1"
     repeat_region   13105. .13137
                     /rpt_family="(CA)n"
     repeat_region   13184. .13309
                     /rpt_family="(CA)n"
     repeat_region   13338. .13638
                     /rpt_family="L1MA5"
     repeat_region   13645. .13895
                     /rpt_family="ORR1C"
     repeat_region   13897. .13997
                     /rpt_family="(TCTCTG)n"
     repeat_region   14209. .14345
                     /rpt_family="L1MC5"
     repeat_region   14923. .14946
                     /rpt_family="AT_rich"
     repeat_region   complement(15136. .15351)
                     /rpt_family="Lx7"
     repeat_region   15352. .15429
                     /rpt_family="MTE"
     repeat_region   15471. .16402
                     /rpt_family="L1_MM"
     repeat_region   complement(16461. .16625)
                     /rpt_family="RMER19B"
     repeat_region   complement(16633. .16943)
                     /rpt_family="RMER19B"
     repeat_region   16949. .17116
                     /rpt_family="L1MA5"
     repeat_region   complement(17295. .17503)
                     /rpt_family="B3"
     repeat_region   complement(17694. .17737)
                     /rpt_family="B3A"
     repeat_region   17738. .17777
                     /rpt_family="(TC)n"
     repeat_region   17777. .17796
                     /rpt_family="(TG)n"
     repeat_region   complement(17778. .17795)
                     /rpt_family="B3A"
     repeat_region   17796. .17825
                     /rpt_family="(GA)n"
     repeat_region   complement(17826. .17967)
```

```
                     /rpt_family="B3A"
     repeat_region   complement(17986. .18121)
                     /rpt_family="MIR"
     repeat_region   complement(18501. .18913)
                     /rpt_family="MTB"
     repeat_region   20008. .20277
                     /rpt_family="MLT1C"
     repeat_region   20603. .20855
                     /rpt_family="MLT1G1"
     repeat_region   20891. .21061

Alignment Scores:
Pred. No.:              20.4          Length:          191497
Score:                  78.00         Matches:         11
Percent Similarity:     100.00%       Conservative:    1
Best Local Similarity:  91.67%        Mismatches:      0
Query Match:            92.86%        Indels:          0
DB:                     9             Gaps:            0

US-09-858-332G-15 (1-12) x AC116420 (1-191497)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              :::||||||||||||||||||||||||||||||||
Db      84863 TATAACCATAACCATAACCATAACCATAACCATAAC 84898

RESULT 170
AC099096
LOCUS       AC099096               215707 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-151I14, *** SEQUENCING IN PROGRESS
            ***, 2 unordered pieces.
ACCESSION   AC099096
VERSION     AC099096.5  GI:30522551
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 215707)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 215707)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-NOV-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 215707)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:22855476.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GGPC
                Center clone name: CH230-151I14
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 208522 bases at least Q40
                Consensus quality: 210852 bases at least Q30
                Consensus quality: 212157 bases at least Q20
                Estimated insert size: 219816; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1   214255: contig of 214255 bp in length
            *    214256   214355: gap of unknown length
            *    214356   215707: contig of 1352 bp in length.
FEATURES                Location/Qualifiers
     source             1. .215707
```

```
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-151I14"
     gap             214256. .214355
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:                22.8          Length:          215707
Score:                    78.00         Matches:         11
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              92.86%        Indels:          0
DB:                       14            Gaps:            0

US-09-858-332G-15 (1-12) x AC099096 (1-215707)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              ||||||||||||||||||||||||||||||||||
Db      89713 CATAACCATAACCATAACCATAACCATAACCAT 89745

RESULT 171
AC098096/c
LOCUS       AC098096               236844 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-168J11, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION   AC098096
VERSION     AC098096.7  GI:30522884
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 236844)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 236844)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 236844)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:23264385.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GGGZ
                Center clone name: CH230-168J11
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 216934 bases at least Q40
                Consensus quality: 220606 bases at least Q30
                Consensus quality: 222998 bases at least Q20
                Estimated insert size: 225975; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   235497: contig of 235497 bp in length
            *  235498   235597: gap of unknown length
            *  235598   236844: contig of 1247 bp in length.
FEATURES                Location/Qualifiers
     source             1. .236844
                        /organism="Rattus norvegicus"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:10116"
```

```
                     /clone="CH230-168J11"
     misc_feature    1. .1141
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    4377. .6479
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    complement(7116. .11456)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH313114"
     misc_feature    227537. .228402
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH313112"
     misc_feature    228402. .229451
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    232108. .235497
                     /note="wgs_end_extension
                     clone_end:T7"
     gap             235498. .235597
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              25              Length:         236844
Score:                  78.00           Matches:        11
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  91.67%          Mismatches:     0
Query Match:            92.86%          Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-15 (1-12) x AC098096 (1-236844)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              :::||||||||||||||||||||||||||||||||
Db      43671 TATAATCATAATCATAATCATAATCATAATCATAAT 43636

RESULT 172
AC118186
LOCUS       AC118186               249236 bp    DNA     linear   HTG 12-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-229B19, WORKING DRAFT SEQUENCE.
ACCESSION   AC118186
VERSION     AC118186.5  GI:23908087
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 249236)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 249236)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-APR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 249236)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Oct 12, 2002 this sequence version replaced gi:21903154.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GWIQ
                Center clone name: CH230-229B19
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 229451 bases at least Q40
                Consensus quality: 231760 bases at least Q30
                Consensus quality: 233324 bases at least Q20
                Estimated insert size: 237494; sum-of-contigs estimation
                Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
            --------------
```

```
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   249236: contig of 249236 bp in length.
FEATURES             Location/Qualifiers
     source          1. .249236
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-229B19"
     misc_feature    complement(1537. .2418)
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:RWBOC10TJ"
     misc_feature    248385. .249233
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:RWBOC10TV"
ORIGIN

Alignment Scores:
Pred. No.:              26.2          Length:        249236
Score:                  78.00         Matches:       11
Percent Similarity:     100.00%       Conservative:  1
Best Local Similarity:  91.67%        Mismatches:    0
Query Match:            92.86%        Indels:        0
DB:                     14            Gaps:          0

US-09-858-332G-15 (1-12) x AC118186 (1-249236)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||:::|||||||||
Db      98069 CACAATCACAATCACAATCACAATTACAATCACAAC 98104

RESULT 173
AX770903
LOCUS       AX770903              249878 bp    DNA     linear   PAT 02-JUL-2003
DEFINITION  Sequence 34 from Patent WO02094867.
ACCESSION   AX770903
VERSION     AX770903.1  GI:32438067
KEYWORDS    .
SOURCE      Photorhabdus luminescens
  ORGANISM  Photorhabdus luminescens
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Photorhabdus.
REFERENCE   1
  AUTHORS   Duchaud,E., Taourit,S., Glaser,P., Frangeul,L., Kunst,F.,
            Danchin,A. and Buchrieser,C.
  TITLE     Sequence of the Photorhabdus luminescens strain TT01 genome and
            uses
  JOURNAL   Patent: WO 02094867-A 34 28-NOV-2002;
            INSTITUT PASTEUR (FR) ; CENTRE NATIONAL DE LA RECHERCHE
            SCIENTIFIQUE (CNRS) (FR)
FEATURES             Location/Qualifiers
     source          1. .249878
                     /organism="Photorhabdus luminescens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:29488"
ORIGIN

Alignment Scores:
Pred. No.:              26.3          Length:        249878
Score:                  78.00         Matches:       11
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            92.86%        Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-15 (1-12) x AX770903 (1-249878)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 11
              ||||||||||||||||||||||||||||||||||||
Db      88838 CACAATCACAATCACAATCACAATCACAATCAC 88870

RESULT 174
AC125668
LOCUS       AC125668              254123 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-21G7, WORKING DRAFT SEQUENCE, 3
            unordered pieces.
ACCESSION   AC125668
VERSION     AC125668.5  GI:30522621
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 254123)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
```

```
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 254123)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 254123)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:23665094.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GGNO
                Center clone name: CH230-21G7
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 240542 bases at least Q40
                Consensus quality: 242651 bases at least Q30
                Consensus quality: 244405 bases at least Q20
                Estimated insert size: 250228; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   250665: contig of 250665 bp in length
            *  250666   250765: gap of unknown length
            *  250766   251914: contig of 1149 bp in length
            *  251915   252014: gap of unknown length
            *  252015   254123: contig of 2109 bp in length.
FEATURES             Location/Qualifiers
     source          1. .254123
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-21G7"
     misc_feature    1. .1679
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    3739. .4568
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH270888"
     misc_feature    196333. .197334
                     /note="wgs_contig"
     misc_feature    247945. .250665
                     /note="wgs_contig"
     gap             250666. .250765
                     /estimated_length=unknown
     gap             251915. .252014
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              26.7          Length:          254123
Score:                  78.00         Matches:         11
Percent Similarity:     100.00%       Conservative:    1
Best Local Similarity:  91.67%        Mismatches:      0
Query Match:            92.86%        Indels:          0
DB:                     14            Gaps:            0

US-09-858-332G-15 (1-12) x AC125668 (1-254123)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||:::||||||||||
Db     239604 CACAATCACAATCACAATCACAATTACAATCACAAC 239639

RESULT 175
BX571864/c
LOCUS       BX571864               349071 bp    DNA     linear   BCT 17-APR-2005
DEFINITION  Photorhabdus luminescens subsp. laumondii TTO1 complete genome;
            segment 6/17.
ACCESSION   BX571864 BX470251
VERSION     BX571864.1  GI:36784828
KEYWORDS    complete genome.
SOURCE      Photorhabdus luminescens subsp. laumondii TTO1
  ORGANISM  Photorhabdus luminescens subsp. laumondii TTO1
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Photorhabdus.
REFERENCE   1
  AUTHORS   Duchaud,E., Rusniok,C., Frangeul,L., Buchrieser,C., Taourit,S.,
            Bocs,S., Boursaux-Eude,C., Chandler,M., Dassa,E., Derose,R.,
            Derzelle,S., Freyssinet,G., Gaudriault,S., Givaudan,A., Glaser,P.,
            Medigue,C., Lanois,A., Powell,K., Siguier,P., Wingate,V.,
            Zouine,M., Boemare,N., Danchin,A. and Kunst,F.
  TITLE     Complete genome sequence of the entomapathogenic bacterium
            Photorhabdus luminescens
  JOURNAL   Nat. Biotechnol. 11 (1) (2003) In press
REFERENCE   2
  AUTHORS   Duchaud,E., Frangeul,L., Rusniok,C. and Kunst,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-APR-2003) L. Frangeul, Institut Pasteur, Genopole, 25
            rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail:
            lfrangeu@pasteur.fr, fkunst@pasteur.fr
FEATURES             Location/Qualifiers
     source          1. .349071
                     /organism="Photorhabdus luminescens subsp. laumondii TTO1"
                     /mol_type="genomic DNA"
                     /strain="TTO1"
                     /db_xref="taxon:243265"
     gene            773. .2351
                     /gene="cydA"
                     /locus_tag="plu1449"
     RBS             773. .778
                     /gene="cydA"
                     /locus_tag="plu1449"
     CDS             783. .2351
                     /gene="cydA"
                     /locus_tag="plu1449"
                     /codon_start=1
                     /transl_table=11
                     /product="Cytochrome D ubiquinol oxidase subunit I
                     (Cytochrome BD-I oxidase subunit I)"
                     /protein_id="CAE13742.1"
                     /db_xref="GI:36784829"
```

```
                /db_xref="GOA:Q7N6U2"
                /db_xref="InterPro:IPR002585"
                /db_xref="UniProt/TrEMBL:Q7N6U2"
                /translation="MFDIVELSRLQFALTAMYHFLFVPLTLGMAFLLAIMETVYVLSG
                KQIYKDMTKFWGKLFGINFALGVATGLTMEFQFGTNWSYFSHYVGDIFGAPLAIEGLM
                AFFLESTFVGLFFFGWDRLSKTQHLAVTWLVALGSNFSALWILVANGWMQNPIASDFN
                FETMRMEMVSFAELVLNPVAQVKFVHTVAAGYVTGAIFVLGVSSYYLLKGRDIPFAKR
                SFAIASSFGIAAVLSVIVLGDESGYEMGDVQKTKLAAIEAEWETQPPPAAFTLIGIPD
                QDKMENKYAIQIPYLMGLIATRSVDTPVIGLRDLMDLHEQRIRNGIKAYHLLEELRAG
                NTDPTVRTAFNENKQDLGYGMLLKRYTSKVTDASEEQIKAAAKDSIPRVAPLYFAFRI
                MVAAGFLMLLLIGLAFWSVIRGRIGKQKWLLRAALYSIPLPWIAIESGWFVAEYGRQP
                WAIGEVLPTAVASSNLAASDILFSMALICGLYTLFLIAEMFLMFKFARLGPSSLKTGR
                YHFEQKTTSAPEAQ"
gene            2354. .3505
                /gene="cydB"
                /locus_tag="plu1450"
RBS             2354. .2359
                /gene="cydB"
                /locus_tag="plu1450"
CDS             2366. .3505
                /gene="cydB"
                /locus_tag="plu1450"
                /codon_start=1
                /transl_table=11
                /product="Cytochrome D ubiquinol oxidase subunit II
                (Cytochrome BD-I oxidase subunit II)"
                /protein_id="CAE13743.1"
                /db_xref="GI:36784830"
                /db_xref="GOA:Q7N6U1"
                /db_xref="InterPro:IPR003317"
                /db_xref="UniProt/TrEMBL:Q7N6U1"
                /translation="MFDYEVLRFIWWLLIGVLLIGFAVTDGFDMGVGILLRIIGKNDT
                ERRIMINSVAPHWDGNQVWLITAGGALFAAWPMVYAAAFSGFYVAMILVLSALFFRPV
                GFDYRSKLENRKWRNMWDWGIFVGSFVPPLVIGVAFGNLLQGVPFNVDSQLRLFYTGS
                FFGLLNPYGLLAGIISLMMIVTQGATYLQMRTTGELRLRSRAATYLCALVTMVAFLLA
                GVWLIYGIDGYIVTGGLDTVAQSNPLHKEVAHQAGAWLTNFNNYPWLWVLPALGVASP
                LLTMLMTRMNKAGWAFLFSSLTIAGIILACGVTMFPFVMPSITDPNVSLTMWDATSSL
                MTLRVMFVVALIFVPIILGYTIWCYYKMFGRLDKSFIEDNKHSLY"
gene            join(4041. .4046,4051. .4455)
                /locus_tag="plu1451"
RBS             4041. .4046
                /locus_tag="plu1451"
CDS             4051. .4455
                /locus_tag="plu1451"
                /function="Unknown"
                /note="unnamed protein product; Highly similar to unknown
                protein YbgC protein of Escherichia coli"
                /codon_start=1
                /transl_table=11
                /protein_id="CAE13744.1"
                /db_xref="GI:36784831"
                /db_xref="GOA:Q7N6U0"
                /db_xref="InterPro:IPR006683"
                /db_xref="InterPro:IPR006684"
                /db_xref="UniProt/TrEMBL:Q7N6U0"
                /translation="MSNTLFRWPIRIYYEDTDVGGIVYNGRYVVFYERARTEMLRQHG
                IHQQRLLDEQIAFVVRRMVVDYRAPARLDDMLEVQSEITDIRRASLTFIQRIVDGNGA
                MINSAEVLVACVNTSQMKPIALPKSIVAEFKQ"
gene            4439. .4444
                /gene="tolQ"
                /locus_tag="plu1452"
RBS             4439. .4444
                /gene="tolQ"
                /locus_tag="plu1452"
gene            4452. .5147
                /gene="tolQ"
                /locus_tag="plu1452"
CDS             4452. .5147
                /gene="tolQ"
                /locus_tag="plu1452"
                /codon_start=1
                /transl_table=11
                /product="TolQ protein"
                /protein_id="CAE13745.1"
                /db_xref="GI:36784832"
                /db_xref="GOA:Q7N6T9"
                /db_xref="InterPro:IPR002898"
                /db_xref="UniProt/TrEMBL:Q7N6T9"
                /translation="MTDMNILDLFLKAGLLVQLIMLILICFSVASWAIIIQRTKILNA
                ATRESEAFEDKFWSGIELSRLYKESQSRRDSLGGTEQIFYSGFKEFARLHHANNHAPE
                AVITGASRAMRISMNRELESLESHIPFLGTVGSISPYIGLFGTVWGIMHAFIALGAVK
                QATLQMVAPGIAEALIATAIGLFAAIPAVMAYNRLNQRVNKLEQNYDNFMEEFLAILH
                RQAFAAENKAENK"
gene            5150. .5587
                /gene="tolR"
                /locus_tag="plu1453"
RBS             5150. .5155
                /gene="tolR"
                /locus_tag="plu1453"
CDS             5162. .5587
                /gene="tolR"
                /locus_tag="plu1453"
                /codon_start=1
                /transl_table=11
                /product="TolR protein"
                /protein_id="CAE13746.1"
                /db_xref="GI:36784833"
                /db_xref="GOA:Q7N6T8"
                /db_xref="InterPro:IPR003400"
                /db_xref="UniProt/TrEMBL:Q7N6T8"
                /translation="MARTRSRKRELKSEINIVPLLDVLLVLLLIFMATAPIITQSVEV
                DLPDAVDSKTVSNSDNPPVIVEVSGVGQYTMIVNHERLELLPEPQIIVEARAQIEKNP
                KTVFLIGGSKEVPYDEIIKALNMLHQAGVKSVGLMTQPI"
gene            5633. .6703
                /gene="tolA"
                /locus_tag="plu1454"
CDS             5633. .6703
                /gene="tolA"
                /locus_tag="plu1454"
                /codon_start=1
                /transl_table=11
                /product="TolA protein"
                /protein_id="CAE13747.1"
                /db_xref="GI:36784834"
                /db_xref="InterPro:IPR010528"
                /db_xref="UniProt/TrEMBL:Q7N6T7"
                /translation="MGKTNGQNNKLNRAVIVSVVLHIILIGLLIWGSLTRKTEMGGGG
                QDGSVIDAVMVDPNAVVQQYNQQQQQQADAKRAEQLRNKKAEQQAAELREKQAEEQER
                LKAVEEERIKAQQAAEEQKKQAVEAANKAREEQKQAEEAAAKAKAEKEKLIREQAEAQ
                KKAEAEAKKEAEAVAAKQKAAEEAKAKADAEAKAKAKAEADAKAKTEADAKAKAKAEA
                DAKKKAAAQAAKQESEVDDLLGGLTSNAPKQQQGASAAGKGGGKKSGASNADINNYLG
                QIQIAIQNKFYDPSMYRGRTCTLRIKLAPDGLLIDVKSEGGDQALCQAAVAAAKQAKI
                PKPPSKEVYEVFKNAPLDFNPQ"
gene            6905. .8209
                /gene="tolB"
                /locus_tag="plu1455"
RBS             6905. .6910
                /gene="tolB"
                /locus_tag="plu1455"
CDS             6917. .8209
                /gene="tolB"
                /locus_tag="plu1455"
                /codon_start=1
                /transl_table=11
                /product="TolB protein precursor"
                /protein_id="CAE13748.1"
                /db_xref="GI:36784835"
                /db_xref="GOA:Q7N6T6"
                /db_xref="InterPro:IPR007195"
                /db_xref="InterPro:IPR011659"
                /db_xref="UniProt/TrEMBL:Q7N6T6"
                /translation="MRQAFKVVLSILMLWVTVAHAEVRIEITQGVDSARPIGVVPFKW
                TGAGNAPEEIGAIVGADLRNSGKFNPIDISRMPQQPTTTSEVTPAAWSALGIDAVVVG
                QIQPSADGGFLISYQLVDTASAPGTVLAQDQFKVTKQWLRYAAHTASDQVFEKLTGIK
                GAFRTRLAYIVKTNGGKYPYELRVSDYDGYNQFTVHRSPEPLMSPAWSPDGAKLAYVT
                FESGRSALVIQTLASGEVRQVASFPRHNGAPAFSPDGSKLAFALSKSGSLNLYVMDLG
```

```
                     SAQIRQITDGRSNNTEPSWMPDNQTLVYTSDQGGRPQLYKTDINGGAPQRLTWEGTQN
                     QDADVSADGNFVMMVSSESGKQHIAKQDLATGAVQILTDTFLDETPSIAPNGTMVIYS
                     STQGLGTVLQLVSTDGRFKARLPATDGQVKFPAWSPYL"
     gene            8240. .8752
                     /gene="pal"
                     /locus_tag="plu1456"
     RBS             8240. .8245
                     /gene="pal"

Alignment Scores:
Pred. No.:              36.3          Length:         349071
Score:                  78.00         Matches:        11
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            92.86%        Indels:         0
DB:                     1             Gaps:           0

US-09-858-332G-15 (1-12) x BX571864 (1-349071)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              ||||||||||||||||||||||||||||||||
Db     260113 CACAATCACAATCACAATCACAATCACAATCAC 260081

RESULT 176
AR551978
LOCUS       AR551978                921 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 7109 from patent US 6747137.
ACCESSION   AR551978
VERSION     AR551978.1  GI:53945153
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 921)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 7109 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .921
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              0.161         Length:         921
Score:                  77.00         Matches:        11
Percent Similarity:     100.00%       Conservative:   1
Best Local Similarity:  91.67%        Mismatches:     0
Query Match:            91.67%        Indels:         0
DB:                     6             Gaps:           0

US-09-858-332G-15 (1-12) x AR551978 (1-921)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||:::||||||||||||||
Db        478 CATAATCATAACCACAATAACAACCACAACCACAAC 513

RESULT 177
AF111943
LOCUS       AF111943                3379 bp    DNA     linear   INV 04-MAY-1999
DEFINITION  Dictyostelium discoideum AX4 development protein DG1124 (DG1124)
            gene, partial cds.
ACCESSION   AF111943
VERSION     AF111943.1  GI:4731915
KEYWORDS    .
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 3379)
  AUTHORS   Iranfar,N. and Loomis,W.F.
  TITLE     Dictyostelium discoideum developmental gene DG1124. Disruption of
            this gene results in morphological defect
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3379)
  AUTHORS   Iranfar,N. and Loomis,W.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1998) Biology, UCSD, 9500 Gilman Drive, Bonner
            Hall, Room 2321, La Jolla, CA 92093, USA
FEATURES             Location/Qualifiers
     source          1. .3379
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
     gene            <1. .>2384
                     /gene="DG1124"
                     /note="disruption results in morphological defects"
     mRNA            join(<1. .368,479. .>2384)
                     /gene="DG1124"
                     /product="development protein DG1124"
     CDS             join(<1. .368,479. .2384)
                     /gene="DG1124"
                     /codon_start=1
                     /product="development protein DG1124"
                     /protein_id="AAD28550.1"
                     /db_xref="GI:4731916"
                     /translation="IDSTGGEEPQQRIEVNQNIPFLIDIKDFGFEKVDIHKTSFSKQQ
                     QQQQQQQQLNKSTTHIKYNSKVMLYSFYSNIEDPSITKKIKFLVSKEDKDISCIGGTW
                     LPDLDGEDVNDHPQTLINTAIRTTKEYTQIDLSGCKKWYKFMEVHYYRPPTEEQSYYQ
                     EITVIYVPDITDIQPMDVDSLLKVPTVSETTPSAPVGVTTATTKTTEEEEEEDQAPKD
                     HVSKDFDNDETSSNTEQNQPPKSSTTSNTTKPAAAVPTTTAATATTTHGSLHLPKASK
                     FYNVITPSSEGNRYKAMMLSLDGLLDYDESDTHEGTFEASLFSELFYLMLCRDMGTNI
                     LTSIVNYRPTSQESDEITDTTLKRKRDDNHAFRGENSEINTSNNTNTNNNNNNNNNNN
                     GTNQPTDSSSLDGNRKSTEKSPQQDVCGVTDKKVKLDGESSTDNTENGEIIGDNTLTN
                     TNNEPVQIEGELNEGLNNNTNNINNNNNNNNEINNNDNNNNNQNENNNQNDNNNEYNI
                     ESNNYNNNDNSSNNTNTSNTSSSSINSPNLSNKDDSNRNSAENQRSSVTSNASSSTST
                     TIGTQVDESAKVLESFHFFDLNRVNYLREEDVETIIHNLGLFLSKSYVQNVVQKVSTN
                     YTSKQGRVYYQEIVKKLIKPTQQQLSTIPSLPQTHSIVPSQSQPQPQQQPQQQPEQQQQ
                     QQQQEQQQQQQEQELMSIEQPQQPTQQQDSMAIEQQSQELAIEQESNLDQPPKDQDLV
                     QQEPIDQQIQQNITENI"
ORIGIN

Alignment Scores:
Pred. No.:              0.562         Length:         3379
Score:                  77.00         Matches:        11
Percent Similarity:     100.00%       Conservative:   1
Best Local Similarity:  91.67%        Mismatches:     0
Query Match:            91.67%        Indels:         0
DB:                     2             Gaps:           0

US-09-858-332G-15 (1-12) x AF111943 (1-3379)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||:::|||||||||
Db       2101 CACAATCACAACCACAACCACAACAACAACCACAAC 2136

RESULT 178
AC138752/c
LOCUS       AC138752                93305 bp    DNA     linear   HTG 16-JAN-2003
DEFINITION  Homo sapiens chromosome 17 clone RP11-1022D1 map 17, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC138752
VERSION     AC138752.1  GI:27764725
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 93305)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-1022D1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 93305)
```

AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (16-JAN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
---------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
---------- Project Information
Center project name: L29179
Center clone name: 1022_D_1
----------.
* NOTE: This record contains 88 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 980: contig of 980 bp in length
* 981 1080: gap of 100 bp
* 1081 2070: contig of 990 bp in length
* 2071 2170: gap of 100 bp
* 2171 3170: contig of 1000 bp in length
* 3171 3270: gap of 100 bp
* 3271 4199: contig of 929 bp in length
* 4200 4299: gap of 100 bp
* 4300 5255: contig of 956 bp in length
* 5256 5355: gap of 100 bp
* 5356 6327: contig of 972 bp in length
* 6328 6427: gap of 100 bp
* 6428 7427: contig of 1000 bp in length
* 7428 7527: gap of 100 bp
* 7528 8492: contig of 965 bp in length
* 8493 8592: gap of 100 bp
* 8593 9533: contig of 941 bp in length
* 9534 9633: gap of 100 bp
* 9634 10574: contig of 941 bp in length
* 10575 10674: gap of 100 bp
* 10675 11702: contig of 1028 bp in length
* 11703 11802: gap of 100 bp
* 11803 12737: contig of 935 bp in length
* 12738 12837: gap of 100 bp
* 12838 13753: contig of 916 bp in length
* 13754 13853: gap of 100 bp
* 13854 14740: contig of 887 bp in length
* 14741 14840: gap of 100 bp
* 14841 15830: contig of 990 bp in length
* 15831 15930: gap of 100 bp
* 15931 16896: contig of 966 bp in length
* 16897 16996: gap of 100 bp
* 16997 18000: contig of 1004 bp in length
* 18001 18100: gap of 100 bp
* 18101 19095: contig of 995 bp in length
* 19096 19195: gap of 100 bp
* 19196 20150: contig of 955 bp in length
* 20151 20250: gap of 100 bp
* 20251 21195: contig of 945 bp in length
* 21196 21295: gap of 100 bp
* 21296 22256: contig of 961 bp in length
* 22257 22356: gap of 100 bp
* 22357 23344: contig of 988 bp in length
* 23345 23444: gap of 100 bp
* 23445 24431: contig of 987 bp in length
* 24432 24531: gap of 100 bp
* 24532 25509: contig of 978 bp in length
* 25510 25609: gap of 100 bp
* 25610 26553: contig of 944 bp in length
* 26554 26653: gap of 100 bp
* 26654 27652: contig of 999 bp in length
* 27653 27752: gap of 100 bp
* 27753 28731: contig of 979 bp in length
* 28732 28831: gap of 100 bp
* 28832 29855: contig of 1024 bp in length
* 29856 29955: gap of 100 bp
* 29956 30972: contig of 1017 bp in length
* 30973 31072: gap of 100 bp
* 31073 32099: contig of 1027 bp in length
* 32100 32199: gap of 100 bp
* 32200 33181: contig of 982 bp in length
* 33182 33281: gap of 100 bp
* 33282 34184: contig of 903 bp in length
* 34185 34284: gap of 100 bp
* 34285 35285: contig of 1001 bp in length
* 35286 35385: gap of 100 bp
* 35386 36352: contig of 967 bp in length
* 36353 36452: gap of 100 bp
* 36453 37356: contig of 904 bp in length
* 37357 37456: gap of 100 bp
* 37457 38433: contig of 977 bp in length
* 38434 38533: gap of 100 bp
* 38534 39566: contig of 1033 bp in length
* 39567 39666: gap of 100 bp
* 39667 40683: contig of 1017 bp in length
* 40684 40783: gap of 100 bp
* 40784 41753: contig of 970 bp in length
* 41754 41853: gap of 100 bp
* 41854 42846: contig of 993 bp in length
* 42847 42946: gap of 100 bp
* 42947 43928: contig of 982 bp in length
* 43929 44028: gap of 100 bp
* 44029 45008: contig of 980 bp in length
* 45009 45108: gap of 100 bp
* 45109 46080: contig of 972 bp in length
* 46081 46180: gap of 100 bp
* 46181 47141: contig of 961 bp in length
* 47142 47241: gap of 100 bp
* 47242 48168: contig of 927 bp in length
* 48169 48268: gap of 100 bp
* 48269 49219: contig of 951 bp in length
* 49220 49319: gap of 100 bp
* 49320 50247: contig of 928 bp in length
* 50248 50347: gap of 100 bp
* 50348 51273: contig of 926 bp in length
* 51274 51373: gap of 100 bp
* 51374 52361: contig of 988 bp in length
* 52362 52461: gap of 100 bp
* 52462 53447: contig of 986 bp in length
* 53448 53547: gap of 100 bp
* 53548 54542: contig of 995 bp in length
* 54543 54642: gap of 100 bp

```
          *    54643    55605: contig of 963 bp in length
          *    55606    55705: gap of 100 bp
          *    55706    56695: contig of 990 bp in length
          *    56696    56795: gap of 100 bp
          *    56796    57759: contig of 964 bp in length
          *    57760    57859: gap of 100 bp
          *    57860    58846: contig of 987 bp in length
          *    58847    58946: gap of 100 bp
          *    58947    59895: contig of 949 bp in length
          *    59896    59995: gap of 100 bp
          *    59996    60962: contig of 967 bp in length
          *    60963    61062: gap of 100 bp
          *    61063    61872: contig of 810 bp in length
          *    61873    61972: gap of 100 bp
          *    61973    62912: contig of 940 bp in length
          *    62913    63012: gap of 100 bp
          *    63013    63991: contig of 979 bp in length
          *    63992    64091: gap of 100 bp
          *    64092    65065: contig of 974 bp in length
          *    65066    65165: gap of 100 bp
          *    65166    66146: contig of 981 bp in length
          *    66147    66246: gap of 100 bp
          *    66247    67234: contig of 988 bp in length
          *    67235    67334: gap of 100 bp
          *    67335    68312: contig of 978 bp in length
          *    68313    68412: gap of 100 bp
          *    68413    69326: contig of 914 bp in length
          *    69327    69426: gap of 100 bp
          *    69427    70361: contig of 935 bp in length
          *    70362    70461: gap of 100 bp
          *    70462    71441: contig of 980 bp in length
          *    71442    71541: gap of 100 bp
          *    71542    72403: contig of 862 bp in length
          *    72404    72503: gap of 100 bp
          *    72504    73490: contig of 987 bp in length
          *    73491    73590: gap of 100 bp
          *    73591    74587: contig of 997 bp in length
          *    74588    74687: gap of 100 bp

Alignment Scores:
Pred. No.:                13.7          Length:          93305
Score:                    77.00         Matches:         11
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    91.67%        Mismatches:      0
Query Match:              91.67%        Indels:          0
DB:                       14            Gaps:            0

US-09-858-332G-15 (1-12) x AC138752 (1-93305)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              :::||||||||||||||||||||||||||||||||
Db      46663 AATAATCATAATCATAATCATAATCATAATCATAAT 46628

RESULT 179
AC003070/c
LOCUS       AC003070              107217 bp    DNA     linear   PRI 31-DEC-2002
DEFINITION  Homo sapiens chromosome 17, clone CTB-39G8, complete sequence.
ACCESSION   AC003070
VERSION     AC003070.2  GI:27436829
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 107217)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone CTB-39G8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 107217)
  AUTHORS   Birren,B., Fasman,K., McKernan,K., Munro,C., Nusbaum,C.,
            Richardson,P., Lander,E., Baldwin,J., Barna,N., Cantu,C., Chang,A.,
            Cooke,P., Daly,M.J., Devon,K., Dewar,K., DuRette,B., Forrest,C.,
            Gage,D., Gensheimer,S., Geraigery,K., Gilmartin,T., Hagos,B.,
            Halphen,I., Harris,K., Howland,J.C., Huang,J., Hui,L., Jacotot,L.,
            Kirby,A., Lane,M., MacKenzie,J., Marquis,N., McDermott,J.,
            Molla,M., Morrow,J., Nachman,A., Naylor,J., O'Connor,T., Olotu,A.,
            Peterson,K., Roberts,D., Rollins,G., Sarnaik,A., Shiu,P., Shyam,R.,
            Stilwell,J., Stone,C., Strickland,C., Sydney,K., Tang,L.,
            Zemtseva,I. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1997) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 107217)
  AUTHORS   Birren,B., Fasman,K., McKernan,K., Nusbaum,C., Richardson,P.,
            Lander,E., Allen,N., Baker,J., Baldwin,J., Barna,N., Beckerly,R.,
            Boutwell,C., Byrne,S., Cantu,C., Castle,A., Cooke,P., Daly,M.J.,
            Depayre,E., Devon,K., Dewar,K., DuRette,B., Etemadi,S.,
            Ferreira,P., Forrest,C., Gage,D., Gardyna,S., Gensheimer,S.,
            Geraigery,K., Gilmartin,T., Gray,D., Hagos,B., Harris,K.,
            Horton,L., Howland,J.C., Hui,L., Jacotot,L., Linton,L.,
            MacKenzie,J., Marquis,N., McEwan,P., McGurk,A., Meldrim,J.,
            Molla,M., Morris,W., Morrow,J., Nachman,A., Naylor,J., O'Connor,T.,
            Pavlin,B., Peterson,K., Ranganath,S., Riley,R., Roberts,D.,
            Rollins,G., Rossello,R., Roy,A., Shyam,R., Soohoo,S., Stilwell,J.,
            Stone,C., Strickland,C., Sydney,K., Tang,L., Vassiliev,H., Vo,A.,
            Wagner,A., Wheeler,J., Wu,Y., Ye,W.J., Zemtseva,I., Zhao,J. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-FEB-1998) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 107217)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-DEC-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Dec 31, 2002 this sequence version replaced gi:2865212.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L266
                Center clone name: 39_G_8
            --------------

            Only the last 107.2 kilobases of this clone are being submitted.
            The remainder overlaps accession number AC091132 [WICGR project
            L11996].
FEATURES             Location/Qualifiers
     source          1. .107217
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
```

```
                     /map="17"
                     /clone="CTB-39G8"
                     /clone_lib="CITB Human BAC"
repeat_region        1. .273
                     /rpt_family="AluJo"
repeat_region        complement(594. .892)
                     /rpt_family="AluSx"
repeat_region        complement(1170. .1338)
                     /rpt_family="MIR"
repeat_region        1452. .1748
                     /rpt_family="AluSx"
repeat_region        1768. .1898
                     /rpt_family="MER5B"
repeat_region        2889. .2920
                     /rpt_family="(TG)n"
repeat_region        complement(5070. .5242)
                     /rpt_family="L3"
repeat_region        complement(5722. .6117)
                     /rpt_family="L2"
repeat_region        6271. .6506
                     /rpt_family="L2"
repeat_region        6903. .7213
                     /rpt_family="AluSx"
repeat_region        complement(7342. .7558)
                     /rpt_family="L2"
repeat_region        7960. .8055
                     /rpt_family="AluSg/x"
repeat_region        8176. .8204
                     /rpt_family="AT_rich"
repeat_region        complement(8227. .8373)
                     /rpt_family="MIR"
repeat_region        complement(8504. .8597)
                     /rpt_family="L1ME4A"
repeat_region        complement(10738. .10924)
                     /rpt_family="MIR"
repeat_region        complement(11137. .11438)
                     /rpt_family="AluSx"
repeat_region        12061. .12182
                     /rpt_family="L3"
repeat_region        13277. .13347
                     /rpt_family="MER117"
repeat_region        complement(13348. .13656)
                     /rpt_family="AluJo"
repeat_region        13657. .13741
                     /rpt_family="MER117"
repeat_region        13761. .13804
                     /rpt_family="(TTA)n"
repeat_region        complement(13806. .14083)
                     /rpt_family="AluSx"
repeat_region        complement(14531. .14831)
                     /rpt_family="AluJo"
repeat_region        complement(15092. .15405)
                     /rpt_family="AluY"
repeat_region        complement(15960. .16262)
                     /rpt_family="AluSg"
repeat_region        16394. .16495
                     /rpt_family="MIR"
repeat_region        17043. .17379
                     /rpt_family="L3"
repeat_region        complement(19972. .20042)
                     /rpt_family="MIR"
repeat_region        complement(20497. .20573)
                     /rpt_family="MIR"
repeat_region        complement(20681. .21202)
                     /rpt_family="MER115"
repeat_region        complement(21398. .21651)
                     /rpt_family="L2"
repeat_region        complement(21783. .21917)
                     /rpt_family="MIR3"
repeat_region        21968. .22045
                     /rpt_family="L2"
repeat_region        complement(22390. .22688)
                     /rpt_family="AluJb"
repeat_region        23207. .23427
                     /rpt_family="L2"
repeat_region        complement(23434. .23482)
                     /rpt_family="L1MB8"
repeat_region        complement(23483. .23795)
                     /rpt_family="AluJb"
repeat_region        complement(23796. .23892)
                     /rpt_family="L1MB8"
repeat_region        complement(23893. .24204)
                     /rpt_family="AluSg1"
repeat_region        complement(24205. .24282)
                     /rpt_family="L1MB8"
repeat_region        complement(24283. .24635)
                     /rpt_family="MLT1A2"
repeat_region        complement(24636. .24810)
                     /rpt_family="L1MB8"
repeat_region        complement(25095. .25453)
                     /rpt_family="LTR16A"
repeat_region        25474. .26467
                     /rpt_family="L2"
repeat_region        26482. .26709
                     /rpt_family="MIR"
repeat_region        26662. .27115
                     /rpt_family="L2"
repeat_region        27118. .27412
                     /rpt_family="AluY"
repeat_region        27436. .27621
                     /rpt_family="MIR"
repeat_region        complement(28227. .28771)
                     /rpt_family="L1MD2"
repeat_region        30191. .30430
                     /rpt_family="L2"
repeat_region        30430. .30544
                     /rpt_family="L2"
repeat_region        30545. .30586
                     /rpt_family="(TGAA)n"

Alignment Scores:
Pred. No.:                15.7          Length:          107217
Score:                    77.00         Matches:         11
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    91.67%        Mismatches:      0
Query Match:              91.67%        Indels:          0
DB:                       8             Gaps:            0

US-09-858-332G-15 (1-12) x AC003070 (1-107217)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              :::||||||||||||||||||||||||||||||||
Db      13802 AATAATCATAATCATAATCATAATCATAATCATAAT 13767

RESULT 180
AC118389_4
WPCOMMENT
  Sequence split into 6 fragments  LOCUS AC118389 Accession AC118389
     Fragment Name      Begin     End
     AC118389_0             1     110000
     AC118389_1        100001     210000
     AC118389_2        200001     310000
     AC118389_3        300001     410000
     AC118389_4        400001     510000
     AC118389_5        500001     568550
Continuation (5 of 6) of AC118389 from base 400001 (AC118389 Rattus norvegicus clone CH2

Alignment Scores:
Pred. No.:                16.1          Length:          110000
Score:                    77.00         Matches:         11
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    91.67%        Mismatches:      0
Query Match:              91.67%        Indels:          0
DB:                       14            Gaps:            0

US-09-858-332G-15 (1-12) x AC118389_4 (1-110000)
```

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              :::||||||||||||||||||||||||||||||||
Db      19220 AACAACCACAACCACAACCACAACCACAACCACAAC 19255

RESULT 181
CT009752_5
WPCOMMENT
  Sequence split into 9 fragments  LOCUS CT009752 Accession CT009752
     Fragment Name      Begin     End
     CT009752_0             1    110000
     CT009752_1        100001    210000
     CT009752_2        200001    310000
     CT009752_3        300001    410000
     CT009752_4        400001    510000
     CT009752_5        500001    610000
     CT009752_6        600001    710000
     CT009752_7        700001    810000
     CT009752_8        800001    868454
Continuation (6 of 9) of CT009752 from base 500001 (CT009752 Trypanosoma brucei chromosc

Alignment Scores:
Pred. No.:              16.1          Length:        110000
Score:                  77.00         Matches:       11
Percent Similarity:     100.00%       Conservative:  1
Best Local Similarity:  91.67%        Mismatches:    0
Query Match:            91.67%        Indels:        0
DB:                     14            Gaps:          0

US-09-858-332G-15 (1-12) x CT009752_5 (1-110000)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              :::||||||||||||||||||||||||||||||||
Db      40542 AATAATCATAATCATAATCATAATCATAATCATAAT 40577

RESULT 182
AC146502/c
LOCUS       AC146502              134798 bp    DNA     linear   VRT 26-NOV-2003
DEFINITION  Danio rerio clone CH211-262M17, complete sequence.
ACCESSION   AC146502
VERSION     AC146502.2  GI:38524632
KEYWORDS    HTG.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 134798)
  AUTHORS   Talbot,W.S., Rauch,G.J., Grimwood,J., Dickson,M., Schmutz,J. and
            Myers,R.M.
  TITLE     Genomic studies of vertebrate diversity
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 134798)
  AUTHORS   Talbot,W.S., Rauch,G.J., Grimwood,J., Dickson,M., Schmutz,J. and
            Myers,R.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-AUG-2003) Stanford Human Genome Center, 975
            California Avenue, Palo Alto, CA 94304, USA
REFERENCE   3  (bases 1 to 134798)
  AUTHORS   Talbot,W.S., Rauch,G.J., Grimwood,J., Dickson,M., Schmutz,J. and
            Myers,R.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-NOV-2003) Stanford Human Genome Center, 975
            California Avenue, Palo Alto, CA 94304, USA
COMMENT     On Nov 26, 2003 this sequence version replaced gi:34013548.
            The sequence of the clone was established as a mapping and
            sequencing  collaboration at the Stanford Genome Evolution Center,
            funded by the NIH Centers of Excellence in Genomic Science (CEGS)
            initiative (http://cegs.stanford.edu).  The clone was isolated from
            the BAC library CHORI-211 (http://bacpac.chori.org).

            Quality: Phrap Quality >=40 100% of Sequence;
            Estimated Total Number of Errors is 0.

FEATURES             Location/Qualifiers
     source          1. .134798
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-262M17"
ORIGIN

Alignment Scores:
Pred. No.:              19.6          Length:        134798
Score:                  77.00         Matches:       11
Percent Similarity:     100.00%       Conservative:  1
Best Local Similarity:  91.67%        Mismatches:    0
Query Match:            91.67%        Indels:        0
DB:                     5             Gaps:          0

US-09-858-332G-15 (1-12) x AC146502 (1-134798)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              :::||||||||||||||||||||||||||||||||
Db      84066 AATAATCATAATCATAATCATAATCATAATCATAAT 84031

RESULT 183
AC079238
LOCUS       AC079238              142462 bp    DNA     linear   PRI 09-JAN-2002
DEFINITION  Homo sapiens BAC clone RP11-719L21 from 4, complete sequence.
ACCESSION   AC079238
VERSION     AC079238.7  GI:16756377
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 142462)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
   PUBMED   9847074
REFERENCE   2  (bases 1 to 142462)
  AUTHORS   Waligorski,J. and Abbott,A.
  TITLE     The sequence of Homo sapiens BAC clone RP11-719L21
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 142462)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 142462)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-NOV-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 142462)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JAN-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   6  (bases 1 to 142462)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Nov 7, 2001 this sequence version replaced gi:16303576.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
```

Center project name: H_NH0719L21

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)
VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-563E2, 2000 bp overlap. Actual start of this clone is at base position 144522 of RP11-563E2; actual end is at base position 142462 of RP11-719L21.

FEATURES	Location/Qualifiers
source	1. .142462
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="4"
	/map="4"
	/clone="RP11-719L21"
	/clone_lib="RPCI-11"
repeat_region	149. .184
	/rpt_family="AT_rich"
repeat_region	390. .870
	/rpt_family="ERV1"
repeat_region	1057. .1197
	/rpt_family="MER103"
repeat_region	1330. .1631
	/rpt_family="Alu"
misc_feature	1599. .1616
	/note="similar to Homo sapiens EST AL522812 (NID:g12786305)"
repeat_region	1610. .1654
	/rpt_family="(CAAA)n"
repeat_region	1734. .1842
	/rpt_family="ERV1"
repeat_region	1843. .3865
	/rpt_family="L1"
repeat_region	3855. .3875
	/rpt_family="AT_rich"
repeat_region	3887. .4224
	/rpt_family="ERV1"
repeat_region	4225. .4505
	/rpt_family="Alu"
repeat_region	4506. .4586
	/rpt_family="ERV1"
repeat_region	4896. .5096
	/rpt_family="MIR"
repeat_region	5146. .5292
	/rpt_family="ERV1"
repeat_region	5283. .5934
	/rpt_family="ERV1"
repeat_region	6157. .6188
	/rpt_family="L2"
repeat_region	6189. .6214
	/rpt_family="(T)n"
repeat_region	6497. .6544
	/rpt_family="L2"
repeat_region	6612. .7007
	/rpt_family="MaLR"
repeat_region	7147. .7206
	/rpt_family="GA-rich"
repeat_region	7263. .7308
	/rpt_family="(TTTA)n"
repeat_region	7282. .7447
	/rpt_family="Alu"
repeat_region	7567. .7704
	/rpt_family="Alu"
repeat_region	7731. .7861
	/rpt_family="L2"
repeat_region	8196. .8496
	/rpt_family="Alu"
repeat_region	8469. .8551
	/rpt_family="GA-rich"
repeat_region	8546. .8723
	/rpt_family="ERV1"
repeat_region	8866. .9371
	/rpt_family="L1"
repeat_region	10399. .10836
	/rpt_family="ERV1"
repeat_region	11121. .11185
	/rpt_family="(TATATG)n"
repeat_region	11284. .11350
	/rpt_family="(TA)n"
repeat_region	11387. .11550
	/rpt_family="L1"
repeat_region	11598. .11706
	/rpt_family="MIR"
repeat_region	13097. .13448
	/rpt_family="MaLR"
repeat_region	15021. .15367
	/rpt_family="MaLR"
repeat_region	16385. .16409
	/rpt_family="(GA)n"
repeat_region	16447. .16636
	/rpt_family="MIR"
repeat_region	17088. .17395
	/rpt_family="L1"
repeat_region	17496. .17781
	/rpt_family="Alu"
repeat_region	18920. .18970
	/rpt_family="L1"
repeat_region	18963. .19064
	/rpt_family="L1"
repeat_region	19117. .19861
	/rpt_family="L1"
repeat_region	19862. .20156
	/rpt_family="Alu"
repeat_region	20157. .20375
	/rpt_family="L1"
repeat_region	21020. .21361
	/rpt_family="MaLR"
repeat_region	21364. .21463
	/rpt_family="MIR"
repeat_region	22606. .22659
	/rpt_family="(GA)n"
repeat_region	23506. .23673
	/rpt_family="ERV1"
repeat_region	23878. .24068
	/rpt_family="ERV1"
repeat_region	24171. .24220
	/rpt_family="L2"

```
     repeat_region   24250. .24327
                     /rpt_family="L2"
     repeat_region   24483. .24590
                     /rpt_family="L2"
     repeat_region   24846. .25044

Alignment Scores:
Pred. No.:              20.6          Length:         142462
Score:                  77.00         Matches:        11
Percent Similarity:     100.00%       Conservative:   1
Best Local Similarity:  91.67%        Mismatches:     0
Query Match:            91.67%        Indels:         0
DB:                     8             Gaps:           0

US-09-858-332G-15 (1-12) x AC079238 (1-142462)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              :::|||||||||||||||||||||||||||||||||
Db      71581 AATAATCATAATCATAATCATAATCATAATCATAAT 71616

RESULT 184
AC117076/c
LOCUS       AC117076              146570 bp    DNA     linear   INV 21-FEB-2004
DEFINITION  Dictyostelium discoideum chromosome 2 map 3323568-3470138 strain
            AX4, complete sequence.
ACCESSION   AC117076 AC115610
VERSION     AC117076.3  GI:42733926
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 146570)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 146570)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-APR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 146570)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 146570)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   5  (bases 1 to 146570)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2004) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On or before Feb 21, 2004 this sequence version replaced
            gi:28828883, gi:28829496.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES            Location/Qualifiers
     source         1. .146570
                    /organism="Dictyostelium discoideum"
                    /mol_type="genomic DNA"
                    /strain="AX4"
                    /db_xref="taxon:44689"
                    /chromosome="2"
                    /map="3323568-3470138"
     CDS            complement(join(<63. .86,436. .603,692. .704))
                    /note="GeneID exon scores (in order of location ranges):
                    1.46, 2.66, -0.45 - GSCJ_ID dd_03180"
                    /codon_start=1
                    /product="hypothetical protein"
                    /protein_id="AAS38882.1"
                    /db_xref="GI:42733988"
                    /translation="MTILGLISSISNVKSISKSNNLSSLSNSLSSLQSMNSIERRGGI
                    SISNPGFLGGIGRNPGESMVCDSG"
     CDS            complement(join(1507. .1763,1872. .1884))
                    /note="GeneID exon scores (in order of location ranges):
                    19.50, 3.22 - GSCJ_ID dd_01276"
                    /codon_start=1
                    /product="hypothetical protein"
                    /protein_id="AAS38835.1"
                    /db_xref="GI:42733927"
                    /translation="MTILGSISSIGNVKSITKSNNLSSLSNSSSSTQSMNSIQCGGGC
                    GNGGLLGGVGGLVGGVLVGTGVIIGSVLHGVGSILTGGSNNCGCN"
     CDS            join(2611. .2623,2691. .2947)
                    /note="GeneID exon scores (in order of location ranges):
                    3.18, 23.32 - GSCJ_ID dd_01275"
                    /codon_start=1
                    /product="hypothetical protein"
                    /protein_id="AAS38836.1"
                    /db_xref="GI:42733928"
                    /translation="MTILGSISSIGNVKLISKSNNLSSLSNSSSSLQSMNSIQCGGGC
                    GNGGLLGGVGGLVGGVLVGTGVIVGSVLHGVGSILTGGSNNCGCN"
     CDS            complement(join(3484. .3740,3824. .3836))
                    /note="GeneID exon scores (in order of location ranges):
                    25.27, 0.79 - GSCJ_ID dd_01274"
                    /codon_start=1
                    /product="hypothetical protein"
                    /protein_id="AAS38837.1"
                    /db_xref="GI:42733929"
                    /translation="MTILGSISSIGNVKSISKSNNLSSLSNSSSSLQSMNSIQCGGGC
                    GNGGLLGAVGGLVGGVLTGTGVIVGSVLHGVGSILTGGPNNCGCN"
     CDS            join(4219. .4228,4659. .4915)
                    /note="GeneID exon scores (in order of location ranges):
                    -7.18, 24.38 - GSCJ_ID dd_03039"
                    /codon_start=1
                    /product="hypothetical protein"
                    /protein_id="AAS38838.1"
                    /db_xref="GI:42733930"
                    /translation="MVKCSISSIGNVKSISKSNNLSSLSNSSSSLQSMNSIQCGGGCG
                    NGGLLGGVGGLVGGVLVGTGVIVGSVLHGVGSILTGGSNNCGCN"
     CDS            complement(join(5405. .5655,5746. .5758))
                    /note="GeneID exon scores (in order of location ranges):
                    25.79, -3.80 - GSCJ_ID dd_03235"
                    /codon_start=1
                    /product="hypothetical protein"
                    /protein_id="AAS38839.1"
                    /db_xref="GI:42733931"
                    /translation="MTILASISSIGNVKSISKSNNVSLSNSSSSLQSMNSIQCGGCGN
                    GGLLGAVGGLVGGVLTGTGVIVGSVLHGVGSILTGGSNNCGCN"
     CDS            join(6487. .6499,6572. .6825)
                    /note="GeneID exon scores (in order of location ranges):
                    3.30, 26.87 - GSCJ_ID dd_03141"
                    /codon_start=1
                    /product="hypothetical protein"
                    /protein_id="AAS38840.1"
                    /db_xref="GI:42733932"
                    /translation="MTILASISSIGNVKSISKSNNLSSLSNSSSSLQSMNSIQCGGCG
                    NGGLLGAVGGLVGGVLTGTGVIVGSVLHGVGSILTGGSNNCGCN"
     CDS            join(7061. .7150,7227. .8279,8516. .9178)
                    /note="GeneID exon scores (in order of location ranges):
                    -0.28, 67.69, 40.02 - GSCJ_ID dd_03341"
                    /codon_start=1
                    /product="similar to Dictyostelium discoideum (Slime
```

mold). MkpA protein"
/protein_id="AAS38841.1"
/db_xref="GI:42733933"
/translation="MDTTKNNDEFDISYFLTCSIFGFILWILTEQILKYYNKTNKNNK
YNLPKGPSFLKWFINYLFNFYDLKLSNNKEEDNNNNNNKSNNSLSQEELIEDTSENTV
LKWFNQLNSDNYSVSFFGRPMIFTRDTTISKYILSSNNIDNYTKPPDSSGVLIRLAQN
SILMSEGDQWRYHRSIINQPFSSKNVKLMIPTIITTINKLINHLNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTIIIDIHSYCTKLTFDII
GKLSIGYDFNSIESSDNDNDNNDDDDISKQFDFILNEMIRPIRRFSSYLPLYNDIKLF
KFLNELESIIKGAINSRSLITDNNNNKTYKKNFLLDNLLDDNVKEKDIIGNINTFLLA
GHETSANLLTFIFYLLSTHNNVQNDLYNHLIENQKKKINKDNKFTEEDEDYQSIEFLD
WVIYETLRLFPPAPMIGRTSKNDDILKSGNNNNNNNNNISIPSETLILISVYAIHRDP
KLWKDPNIFNPYRWKNIENINNRSDFIPFSSGGRVCVGQKFSIVEARIIISKLILNFE
LSFNNLKSKPFKIYQRATLTPKYPVFLNFKKRENK"
CDS complement(9415. .11361)
/note="GeneID exon scores (in order of location ranges):
133.26 - GSCJ_ID dd_03256"
/codon_start=1
/product="similar to Dictyostelium discoideum (Slime
mold). CIGB protein (Fragment)"
/protein_id="AAS38842.1"
/db_xref="GI:42733934"
/translation="MMEPINYKNISFNDIKLDSNCNSSSSSSSSSIGSNRSGSGGGGA
ISRKSFDCSTSNGISEFGGDDGGCGGGCGGDMGNNINISINKRHNSVNSISSNKNINC
LYHDLNDLITYCNDCERMICVQCIVSNDHCGHKFTEISKPFASSLLEKFQKQQFPNLL
GYRESDKHIISALDYCCNDVERQHIESVKQVKQHFQKLIESIQSIEKDIIEKLQSTVE
WNKELQNSIKQQFERDSKKIDSIISKHHSKINGYQIDQIYDFETKQHIEIIKDSDQCT
PLLKRRLEILNNNSSSKIKFKNDLLTFNMEIFDEIKEKVLNFTKIENVEYKDLKSTKQ
HLMLNKKIQPYNIDYYLCHGEPIPLGETSVALDSKYSSKSICIQSNKSIDRLLLCNGF
KLSLSEFKIPDSIKTLRIGDIKYSWDANSIPSSVTELHLCDGFNHILSSGMIPNTVTK
LYLYDIKQPLTMGSIPSSVNELHMCDGFSLQLAPFVLPNSITSLSIGEIKYSILPEFI
PSSVKKIRLYSSFNHPIIPQNISESVETLVLFNIKLNLENKVIPKSIKELHLCDGFEQ
QIQPFSIENGGKLEKLVLCDIKEKLLKNSIPNSVKKLIFLGYNHQLENGILNDSLELI
YFSKSFTHPIDNNLIPNNIEIIKK"
CDS complement(join(12693. .12990,13133. .13227))
/note="GeneID exon scores (in order of location ranges):
18.88, 1.20 - GSCJ_ID dd_02598"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAO52031.2"
/db_xref="GI:42733935"
/translation="MSLNNFSPEEKQNIVRSIFAEEEIVYNVFFERGDDSFKNLKLSE
IKAILSILKSSRSMAYPLSGKKQVLVDLLTSIGFVGPINTSPQPQPPQPPPSPTPPGL
FQEELQPHSFLELDLGITVELDNLLGEN"
CDS join(13881. .13926,14100. .14230,14342. .14589,14703. .15129)
/note="GeneID exon scores (in order of location ranges):
6.31, 2.00, 17.53, 44.25 - GSCJ_ID dd_02599"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAO52033.2"
/db_xref="GI:42733936"
/translation="MTLVEYSETLKKSLEEFCKQDKKVVTPELNSIIENISKTGITCY
PWNILKELFYFKLNEILDIFEKEYIATSSTTTTSTTTTTTTNTSSSSTNSPQSPQSPQS
PQLNVNIKDEEMQKIKNQVNLSGLIKMKKEFLDSFKESKLSPFTVQRLCELIINYKMY
TSFSKYLCAVEKMLNVSTLPHLTPEEVIEFNKNQNSGSRLNSSTNTTTTTTTTTTTTT
PTEKGNNDQPNLDSFAFSSSFTTSFPSDTTSSTSTSLSTDQEMKDLDVEQQKDDTDNT
ANENESK"
CDS complement(15517. .15972)
/note="GeneID exon scores (in order of location ranges):
30.20 - GSCJ_ID dd_02600"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAO52034.2"
/db_xref="GI:42733937"
/translation="MDPGRSHIETIIESSTHKFILFDEPTEETIPYFKDLMKKNSCIN
IVRCCEINYDASLFEGVKIHELCFKDGNVPPKDIIERWLEILKQAFIENGKQKTTVGI
HCIAGLGRTPLLVCIALIEDGMKPLQAVEFVRSKRKNAINSPQIKFLKE"
CDS complement(join(17629. .18615,18760. .18992,19147. .19282))
/note="GeneID exon scores (in order of location ranges):
132.77, 35.41, 19.03 - GSCJ_ID dd_02602"
/codon_start=1
/product="similar to seryl-trna synthetase, cytoplasmic
(EC 6.1.1.11) [Schizosaccharomyces pombe]"

/protein_id="AAS38843.1"
/db_xref="GI:42733938"
/translation="MGLDIVLFRADKGGNPDLIRQSQKVRYANVDAVQEVIDLDKVVK

Alignment Scores:

Pred. No.:	21.2	Length:	146570
Score:	77.00	Matches:	11
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	91.67%	Mismatches:	0
Query Match:	91.67%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-15 (1-12) x AC117076 (1-146570)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||:::|||
Db       6207 CACAATCACAATCACAATCACAATCACAATAACAAT 6172
```

RESULT 185
AC141999
LOCUS AC141999 160634 bp DNA linear HTG 21-MAR-2003
DEFINITION Rattus norvegicus clone CH230-270C13, *** SEQUENCING IN PROGRESS ***, 61 unordered pieces.
ACCESSION AC141999
VERSION AC141999.1 GI:29135470
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 160634)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., Mcleod,M., Mcneill,T., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S.,

```
            Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D.,
            Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J.,
            Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R.,
            Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S.,
            Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X.,
            Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R.,
            Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 160634)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
              Center project name: GRBY
              Center clone name: CH230-270C13
            -------------- Summary Statistics
              Sequencing vector: Plasmid;
              Chemistry: Dye-terminator Big Dye: 100% of reads
              Assembly program: Phrap; version 0.990329
              Consensus quality: 126970 bases at least Q40
              Consensus quality: 134664 bases at least Q30
              Consensus quality: 140158 bases at least Q20
              Estimated insert size: 126396; sum-of-contigs estimation
              Quality coverage: 2x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 61 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1002: contig of 1002 bp in length
            *    1003     1102: gap of unknown length
            *    1103     2104: contig of 1002 bp in length
            *    2105     2204: gap of unknown length
            *    2205     3741: contig of 1537 bp in length
            *    3742     3841: gap of unknown length
            *    3842     5455: contig of 1614 bp in length
            *    5456     5555: gap of unknown length
            *    5556     7319: contig of 1764 bp in length
            *    7320     7419: gap of unknown length
            *    7420     8494: contig of 1075 bp in length
            *    8495     8594: gap of unknown length
            *    8595    10312: contig of 1718 bp in length
            *   10313    10412: gap of unknown length
            *   10413    11608: contig of 1196 bp in length
            *   11609    11708: gap of unknown length
            *   11709    12796: contig of 1088 bp in length
            *   12797    12896: gap of unknown length
            *   12897    14011: contig of 1115 bp in length
            *   14012    14111: gap of unknown length
            *   14112    15372: contig of 1261 bp in length
            *   15373    15472: gap of unknown length
            *   15473    16702: contig of 1230 bp in length
            *   16703    16802: gap of unknown length
            *   16803    18540: contig of 1738 bp in length
            *   18541    18640: gap of unknown length
            *   18641    19848: contig of 1208 bp in length
            *   19849    19948: gap of unknown length
            *   19949    21085: contig of 1137 bp in length
            *   21086    21185: gap of unknown length
            *   21186    22538: contig of 1353 bp in length
            *   22539    22638: gap of unknown length
            *   22639    24199: contig of 1561 bp in length
            *   24200    24299: gap of unknown length
            *   24300    25490: contig of 1191 bp in length
            *   25491    25590: gap of unknown length
            *   25591    26810: contig of 1220 bp in length
            *   26811    26910: gap of unknown length
            *   26911    28333: contig of 1423 bp in length
            *   28334    28433: gap of unknown length
            *   28434    29824: contig of 1391 bp in length
            *   29825    29924: gap of unknown length
            *   29925    31363: contig of 1439 bp in length
            *   31364    31463: gap of unknown length
            *   31464    33704: contig of 2241 bp in length
            *   33705    33804: gap of unknown length
            *   33805    35873: contig of 2069 bp in length
            *   35874    35973: gap of unknown length
            *   35974    38042: contig of 2069 bp in length
            *   38043    38142: gap of unknown length
            *   38143    39581: contig of 1439 bp in length
            *   39582    39681: gap of unknown length
            *   39682    41011: contig of 1330 bp in length
            *   41012    41111: gap of unknown length
            *   41112    42348: contig of 1237 bp in length
            *   42349    42448: gap of unknown length
            *   42449    44108: contig of 1660 bp in length
            *   44109    44208: gap of unknown length
            *   44209    47160: contig of 2952 bp in length
            *   47161    47260: gap of unknown length
            *   47261    49178: contig of 1918 bp in length
            *   49179    49278: gap of unknown length
            *   49279    51273: contig of 1995 bp in length
            *   51274    51373: gap of unknown length
            *   51374    53195: contig of 1822 bp in length
            *   53196    53295: gap of unknown length
            *   53296    55875: contig of 2580 bp in length
            *   55876    55975: gap of unknown length
            *   55976    58197: contig of 2222 bp in length
            *   58198    58297: gap of unknown length
            *   58298    60391: contig of 2094 bp in length
            *   60392    60491: gap of unknown length
            *   60492    61807: contig of 1316 bp in length
            *   61808    61907: gap of unknown length
            *   61908    64495: contig of 2588 bp in length
            *   64496    64595: gap of unknown length
            *   64596    67408: contig of 2813 bp in length
            *   67409    67508: gap of unknown length
            *   67509    69971: contig of 2463 bp in length
            *   69972    70071: gap of unknown length
            *   70072    72989: contig of 2918 bp in length
            *   72990    73089: gap of unknown length
            *   73090    74939: contig of 1850 bp in length
            *   74940    75039: gap of unknown length
            *   75040    76951: contig of 1912 bp in length
            *   76952    77051: gap of unknown length
            *   77052    80255: contig of 3204 bp in length
            *   80256    80355: gap of unknown length
            *   80356    82038: contig of 1683 bp in length
            *   82039    82138: gap of unknown length
            *   82139    85079: contig of 2941 bp in length
            *   85080    85179: gap of unknown length
            *   85180    88512: contig of 3333 bp in length
            *   88513    88612: gap of unknown length
            *   88613    92252: contig of 3640 bp in length
            *   92253    92352: gap of unknown length
            *   92353    94705: contig of 2353 bp in length
            *   94706    94805: gap of unknown length
            *   94806    98359: contig of 3554 bp in length
            *   98360    98459: gap of unknown length
            *   98460   103311: contig of 4852 bp in length
            *  103312   103411: gap of unknown length
            *  103412   107829: contig of 4418 bp in length
```

```
          *   107830   107929: gap of unknown length

Alignment Scores:
Pred. No.:                23.2          Length:         160634
Score:                    77.00         Matches:        11
Percent Similarity:       100.00%       Conservative:   1
Best Local Similarity:    91.67%        Mismatches:     0
Query Match:              91.67%        Indels:         0
DB:                       14            Gaps:           0

US-09-858-332G-15 (1-12) x AC141999 (1-160634)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||:::|||
Db      21226 CACAACCACAACCACAACCACAACCACAACAACAAC 21261

RESULT 186
AC146504/c
LOCUS       AC146504              181948 bp    DNA     linear   VRT 05-MAY-2004
DEFINITION  Danio rerio clone CH211-86L16, complete sequence.
ACCESSION   AC146504
VERSION     AC146504.2  GI:47026470
KEYWORDS    HTG.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 181948)
  AUTHORS   Talbot,W.S., Rauch,G.J., Grimwood,J., Dickson,M., Schmutz,J. and
            Myers,R.M.
  TITLE     Genomic studies of vertebrate diversity
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 181948)
  AUTHORS   Talbot,W.S., Rauch,G.J., Grimwood,J., Dickson,M., Schmutz,J. and
            Myers,R.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-AUG-2003) Stanford Human Genome Center, 975
            California Avenue, Palo Alto, CA 94304, USA
REFERENCE   3  (bases 1 to 181948)
  AUTHORS   Talbot,W.S., Rauch,G.J., Grimwood,J., Dickson,M., Schmutz,J. and
            Myers,R.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAY-2004) Stanford Human Genome Center, 975
            California Avenue, Palo Alto, CA 94304, USA
COMMENT     On May 5, 2004 this sequence version replaced gi:34013550.
            The sequence of the clone was established as a mapping and
            sequencing collaboration at the Stanford Genome Evolution Center,
            funded by the NIH Centers of Excellence in Genomic Science (CEGS)
            initiative (http://cegs.stanford.edu). The clone was isolated from
            the BAC library CHORI-211 (http://bacpac.chori.org).

            Quality: Phrap Quality >=40 100% of Sequence;
            Estimated Total Number of Errors is 0.6.
FEATURES             Location/Qualifiers
     source          1. .181948
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-86L16"
ORIGIN

Alignment Scores:
Pred. No.:                26.1          Length:         181948
Score:                    77.00         Matches:        11
Percent Similarity:       100.00%       Conservative:   1
Best Local Similarity:    91.67%        Mismatches:     0
Query Match:              91.67%        Indels:         0
DB:                       5             Gaps:           0

US-09-858-332G-15 (1-12) x AC146504 (1-181948)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              :::||||||||||||||||||||||||||||||
Db     128275 AATAATCATAATCATAATCATAATCATAATCATAAT 128240

RESULT 187
CR524481/c
LOCUS       CR524481              191642 bp    DNA     linear   VRT 06-AUG-2004
DEFINITION  Zebrafish DNA sequence from clone CH211-81C11 in linkage group 22,
            complete sequence.
ACCESSION   CR524481
VERSION     CR524481.5  GI:51036005
KEYWORDS    HTG.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 191642)
  AUTHORS   Clark,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 6, 2004 this sequence version replaced gi:50838643.
            ------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived
            Zebrafish pUC subclones occasionally display inconsistency over the
            length of mononucleotide A/T runs and conserved TA repeats. Where
            this is found the longest good quality representation will be
            submitted.
            Repeat names beginning 'Dr' were identified by the Recon repeat
            discovery system (Zhirong Bao and Sean Eddy, submitted), and those
            beginning 'drr' were identified by Rick Waterman (Stephen Johnson
            lab, WashU). For further information see
            http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml CH211-81C11
            is from a CHORI-211 BAC library
            VECTOR: pTARBAC2.1.
FEATURES             Location/Qualifiers
     source          1. .191642
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-81C11"
                     /clone_lib="CHORI-211"
ORIGIN

Alignment Scores:
Pred. No.:                27.5          Length:         191642
Score:                    77.00         Matches:        11
Percent Similarity:       100.00%       Conservative:   1
Best Local Similarity:    91.67%        Mismatches:     0
```

```
Query Match:            91.67%        Indels:        0
DB:                     5             Gaps:          0

US-09-858-332G-15 (1-12) x CR524481 (1-191642)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              :::||||||||||||||||||||||||||||||||||
Db     128213 AATAATCATAATCATAATCATAATCATAATCATAAT 128178

RESULT 188
AC132694
LOCUS       AC132694               202314 bp    DNA     linear   HTG 15-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-490C22, WORKING DRAFT SEQUENCE.
ACCESSION   AC132694
VERSION     AC132694.3  GI:25007304
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 202314)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 202314)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 202314)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 15, 2002 this sequence version replaced gi:23908401.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: KCCJ
                Center clone name: CH230-490C22
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 169128 bases at least Q40
                Consensus quality: 170817 bases at least Q30
                Consensus quality: 171921 bases at least Q20
                Estimated insert size: 173670; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *         1   202314: contig of 202314 bp in length.
FEATURES             Location/Qualifiers
     source          1. .202314
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-490C22"
     misc_feature    1. .2170
                     /note="wgs_contig"
     misc_feature    2221. .3503
                     /note="wgs_contig"
     misc_feature    complement(199582. .200007)
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:BZ125806"
     misc_feature    201299. .202314
                     /note="wgs_end_extension
                     clone_end:T7"
ORIGIN
```

```
Alignment Scores:
Pred. No.:                28.9          Length:          202314
Score:                    77.00         Matches:         11
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    91.67%        Mismatches:      0
Query Match:              91.67%        Indels:          0
DB:                       14            Gaps:            0

US-09-858-332G-15 (1-12) x AC132694 (1-202314)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||:::||||||||||||||||
Db     148744 CACAACCACAACCACAACAACAACCACAACCACAAC 148779

RESULT 189
AC123276
LOCUS       AC123276               218242 bp    DNA     linear   HTG 11-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-342L10, *** SEQUENCING IN PROGRESS
            ***, 3 unordered pieces.
ACCESSION   AC123276
VERSION     AC123276.3  GI:23618320
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 218242)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 218242)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 218242)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Oct 9, 2002 this sequence version replaced gi:21671609.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GTVT
                Center clone name: CH230-342L10
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 180118 bases at least Q40
                Consensus quality: 184389 bases at least Q30
                Consensus quality: 187102 bases at least Q20
                Estimated insert size: 184247; sum-of-contigs estimation
                Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1   215792: contig of 215792 bp in length
            *    215793   215892: gap of unknown length
            *    215893   217076: contig of 1184 bp in length
            *    217077   217176: gap of unknown length
            *    217177   218242: contig of 1066 bp in length.
FEATURES                Location/Qualifiers
     source             1. .218242
                        /organism="Rattus norvegicus"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:10116"
                        /clone="CH230-342L10"
     misc_feature       1. .1284
                        /note="wgs_end_extension
                        clone_end:T7"
     misc_feature       4161. .5024
                        /note="clone_boundary
```

```
                     clone_end:T7
                     site:MboI
                     end_sequence:RXAQX65TJ"
     misc_feature    100303. .101682
                     /note="wgs_contig"
     misc_feature    complement(214735. .215604)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:MboI
                     end_sequence:RXAQX65TV"
     gap             215793. .215892
                     /estimated_length=unknown
     gap             217077. .217176
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:                31.1          Length:          218242
Score:                    77.00         Matches:         11
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    91.67%        Mismatches:      0
Query Match:              91.67%        Indels:          0
DB:                       14            Gaps:            0

US-09-858-332G-15 (1-12) x AC123276 (1-218242)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              :::||||||||||||||||||||||||||||||||
Db      20413 AACAACCACAACCACAACCACAACCACAACCACAAT 20448

RESULT 190
AC136813/c
LOCUS       AC136813               226154 bp    DNA     linear   HTG 23-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-102O15, *** SEQUENCING IN PROGRESS
            ***, 2 unordered pieces.
ACCESSION   AC136813
VERSION     AC136813.2  GI:25188343
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 226154)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 226154)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 226154)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 23, 2002 this sequence version replaced gi:24796686.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: KCNT
                Center clone name: CH230-102O15
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 202177 bases at least Q40
                Consensus quality: 206019 bases at least Q30
                Consensus quality: 208572 bases at least Q20
                Estimated insert size: 209740; sum-of-contigs estimation
                Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
```

```
            *        1   224327: contig of 224327 bp in length
            *   224328   224427: gap of unknown length
            *   224428   226154: contig of 1727 bp in length.
FEATURES             Location/Qualifiers
     source          1. .226154
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-102O15"
     misc_feature    1. .1677
                     /note="wgs_contig"
     misc_feature    93221. .95615
                     /note="wgs_contig"
     misc_feature    151021. .152346
                     /note="wgs_contig"
     gap             224328. .224427
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:                32.2          Length:         226154
Score:                    77.00         Matches:        11
Percent Similarity:       100.00%       Conservative:   1
Best Local Similarity:    91.67%        Mismatches:     0
Query Match:              91.67%        Indels:         0
DB:                       14            Gaps:           0

US-09-858-332G-15 (1-12) x AC136813 (1-226154)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              :::|||||||||||||||||||||||||||||||||
Db     144642 AACAACCACAACCACAACCACAACCACAACCACAAC 144607

RESULT 191
AC151871
LOCUS       AC151871               229478 bp    DNA     linear   HTG 09-OCT-2004
DEFINITION  Lemur catta clone LB2-91K12, WORKING DRAFT SEQUENCE, 9 ordered
            pieces.
ACCESSION   AC151871
VERSION     AC151871.1  GI:54019467
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Lemur catta (ring-tailed lemur)
  ORGANISM  Lemur catta
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Strepsirhini;
            Lemuridae; Lemur.
REFERENCE   1  (bases 1 to 229478)
  AUTHORS   Cheng,J.-F., Hamilton,M., Peng,Y., Mukherjee,S., Hosseini,R.,
            Peng,Z., Malinov,I. and Rubin,E.M.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 229478)
  AUTHORS   Cheng,J.-F., Hamilton,M., Peng,Y., Mukherjee,S., Hosseini,R.,
            Peng,Z., Malinov,I. and Rubin,E.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-2004) Genome Sciences, Lawrence Berkeley National
            Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA
COMMENT
            Sequence Produced by Berkeley PGA
            Web site: http://pga.lbl.gov
            Center Code: PGABERK
            Center Project Name: L148-91K12
            Bac Clone Name: LB2-91K12

            This sequence has been compared to sequences of other species
            using Vista (http://www-gsd.lbl.gov/VISTA). The results can be
            viewed at:
            http://pga.lbl.gov/cgi-bin/search_cvcgd?type=n&value=CYP7A1

            The order-orientation of the draft sequence was accomplished by
            using:
            Avid (http://baboon.math.berkeley.edu/mavid) and paired end
            information.

            Funding agent: Programs for Genomic Applications (NHLBI)

            If the Bac Library Name is LB1 to LB4, please see website
            for the description: http://www-gsd.lbl.gov/cheng/BAC.html
            These libraries are available through the BACPAC Resources Center:
            http://www.chori.org/bacpac/libraryres.htm as LBNL-1 to LBNL-4.

            Summary Statistics:
            Sequencing vector: Plasmid; pUC18
            Chemistry: Dye-terminator Big Dye
            Assembly program: Phrap version 0.990329.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1     2062: contig of 2062 bp in length
            *     2063     2162: gap of unknown length
            *     2163     9982: contig of 7820 bp in length
            *     9983    10082: gap of unknown length
            *    10083    25470: contig of 15388 bp in length
            *    25471    25570: gap of unknown length
            *    25571    44902: contig of 19332 bp in length
            *    44903    45002: gap of unknown length
            *    45003    72515: contig of 27513 bp in length
            *    72516    72615: gap of unknown length
            *    72616   101967: contig of 29352 bp in length
            *   101968   102067: gap of unknown length
            *   102068   134324: contig of 32257 bp in length
            *   134325   134424: gap of unknown length
            *   134425   178837: contig of 44413 bp in length
            *   178838   178937: gap of unknown length
            *   178938   229478: contig of 50541 bp in length.
FEATURES             Location/Qualifiers
     source          1. .229478
                     /organism="Lemur catta"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9447"
                     /clone="LB2-91K12"
     gap             2063. .2162
                     /estimated_length=unknown
     gap             9983. .10082
                     /estimated_length=unknown
     gap             25471. .25570
                     /estimated_length=unknown
     gap             44903. .45002
                     /estimated_length=unknown
     gap             72516. .72615
                     /estimated_length=unknown
     gap             101968. .102067
                     /estimated_length=unknown
     gap             134325. .134424
                     /estimated_length=unknown
     gap             178838. .178937
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:                32.7          Length:         229478
Score:                    77.00         Matches:        11
Percent Similarity:       100.00%       Conservative:   1
Best Local Similarity:    91.67%        Mismatches:     0
Query Match:              91.67%        Indels:         0
DB:                       14            Gaps:           0

US-09-858-332G-15 (1-12) x AC151871 (1-229478)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
```

```
                :::||||||||||||||||||||||||||||||||
Db      145643 AACAACCACAACCACAACCACAACCACAACCACAAC 145678

RESULT 192
AC133687
LOCUS       AC133687               233887 bp    DNA     linear   HTG 19-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-201C8, *** SEQUENCING IN PROGRESS
            ***, 2 unordered pieces.
ACCESSION   AC133687
VERSION     AC133687.2  GI:25084488
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 233887)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 233887)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 233887)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 19, 2002 this sequence version replaced gi:22953920.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GMRX
                Center clone name: CH230-201C8
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 211151 bases at least Q40
                Consensus quality: 213294 bases at least Q30
                Consensus quality: 214516 bases at least Q20
                Estimated insert size: 217289; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   232655: contig of 232655 bp in length
            *  232656   232755: gap of unknown length
            *  232756   233887: contig of 1132 bp in length.
FEATURES             Location/Qualifiers
     source          1. .233887
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-201C8"
     misc_feature    1. .1184
                     /note="wgs_contig"
     misc_feature    1235. .4520
                     /note="wgs_contig"
     misc_feature    complement(228591. .229088)
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:BH330304"
     misc_feature    229142. .232655
                     /note="wgs_end_extension
                     clone_end:T7"
     gap             232656. .232755
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              33.3          Length:        233887
```

```
Score:                  77.00       Matches:        11
Percent Similarity:     100.00%     Conservative:   1
Best Local Similarity:  91.67%      Mismatches:     0
Query Match:            91.67%      Indels:         0
DB:                     14          Gaps:           0

US-09-858-332G-15 (1-12) x AC133687 (1-233887)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              :::||||||||||||||||||||||||||||||||
Db     190077 AACAACCACAACCACAACCACAACCACAACCACAAT 190112

RESULT 193
AC128004
LOCUS       AC128004               247726 bp    DNA     linear   HTG 20-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-262L8, *** SEQUENCING IN PROGRESS
            ***, 3 unordered pieces.
ACCESSION   AC128004
VERSION     AC128004.3  GI:25139641
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 247726)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 247726)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 247726)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 20, 2002 this sequence version replaced gi:23665065.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: KAJE
                Center clone name: CH230-262L8
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 193410 bases at least Q40
                Consensus quality: 198067 bases at least Q30
                Consensus quality: 201676 bases at least Q20
                Estimated insert size: 196651; sum-of-contigs estimation
                Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1   245348: contig of 245348 bp in length
            *    245349   245448: gap of unknown length
            *    245449   246558: contig of 1110 bp in length
            *    246559   246658: gap of unknown length
            *    246659   247726: contig of 1068 bp in length.
FEATURES             Location/Qualifiers
     source          1. .247726
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-262L8"
     misc_feature    1. .1135
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    4179. .4957
                     /note="clone_boundary
                     clone_end:T7
                     site:
```

```
                end_sequence:BZ267247"
     misc_feature    205970. .207570
                     /note="wgs_contig"
     misc_feature    208072. .209311
                     /note="wgs_contig"
     gap             245349. .245448
                     /estimated_length=unknown
     gap             246559. .246658
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              35.2          Length:         247726
Score:                  77.00         Matches:        11
Percent Similarity:     100.00%       Conservative:   1
Best Local Similarity:  91.67%        Mismatches:     0
Query Match:            91.67%        Indels:         0
DB:                     14            Gaps:           0

US-09-858-332G-15 (1-12) x AC128004 (1-247726)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||:::||||||||||||||||
Db      70129 CACAACCACAACCACAACAACAACCACAACCACAAC 70164

RESULT 194
AC095505
LOCUS       AC095505               251231 bp    DNA     linear   HTG 09-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-7P19, WORKING DRAFT SEQUENCE, 3
            unordered pieces.
ACCESSION   AC095505
VERSION     AC095505.6  GI:30467578
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 251231)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 251231)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 251231)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 9, 2003 this sequence version replaced gi:24941016.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GCMK
                Center clone name: CH230-7P19
            -------------- Summary Statistics
                Assembly program: Atlas;
                Consensus quality: 235839 bases at least Q40
                Consensus quality: 238363 bases at least Q30
                Consensus quality: 239820 bases at least Q20
                Estimated insert size: 250050; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   244990: contig of 244990 bp in length
            *  244991   245090: gap of unknown length
            *  245091   246835: contig of 1745 bp in length
            *  246836   246935: gap of unknown length
```

```
          *    246936   251231: contig of 4296 bp in length.
FEATURES             Location/Qualifiers
     source          1. .251231
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-7P19"
     misc_feature    1. .1831
                     /note="wgs_contig"
     gap             244991. .245090
                     /estimated_length=unknown
     gap             246836. .246935
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:                35.6          Length:          251231
Score:                    77.00         Matches:         11
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    91.67%        Mismatches:      0
Query Match:              91.67%        Indels:          0
DB:                       14            Gaps:            0

US-09-858-332G-15 (1-12) x AC095505 (1-251231)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              :::||||||||||||||||||||||||||||||||
Db     151103 AACAACCACAACCACAACCACAACCACAACCACAAC 151138

RESULT 195
AC120488
LOCUS       AC120488               253932 bp    DNA     linear   HTG 08-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-120E16, *** SEQUENCING IN PROGRESS
            ***, 6 unordered pieces.
ACCESSION   AC120488
VERSION     AC120488.4  GI:22856140
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 253932)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 253932)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 253932)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 14, 2002 this sequence version replaced gi:21908639.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GXUM
                Center clone name: CH230-120E16
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 233785 bases at least Q40
                Consensus quality: 237431 bases at least Q30
                Consensus quality: 239480 bases at least Q20
                Estimated insert size: 255250; sum-of-contigs estimation
                Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
```

```
            *        1   246663: contig of 246663 bp in length
            *   246664   246763: gap of unknown length
            *   246764   248117: contig of 1354 bp in length
            *   248118   248217: gap of unknown length
            *   248218   249267: contig of 1050 bp in length
            *   249268   249367: gap of unknown length
            *   249368   250411: contig of 1044 bp in length
            *   250412   250511: gap of unknown length
            *   250512   251954: contig of 1443 bp in length
            *   251955   252054: gap of unknown length
            *   252055   253932: contig of 1878 bp in length.
FEATURES             Location/Qualifiers
     source          1. .253932
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-120E16"
     misc_feature    1. .1533
                     /note="wgs_contig"
     misc_feature    245478. .246663
                     /note="wgs_contig"
     gap             246664. .246763
                     /estimated_length=unknown
     gap             248118. .248217
                     /estimated_length=unknown
     gap             249268. .249367
                     /estimated_length=unknown
     gap             250412. .250511
                     /estimated_length=unknown
     gap             251955. .252054
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              36              Length:         253932
Score:                  77.00           Matches:        11
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  91.67%          Mismatches:     0
Query Match:            91.67%          Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-15 (1-12) x AC120488 (1-253932)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              :::||||||||||||||||||||||||||||||||
Db     165010 AATAATCATAATCATAATCATAATCATAATCATAAT 165045

RESULT 196
AC121746
LOCUS       AC121746              265001 bp    DNA     linear   HTG 20-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-109L10, WORKING DRAFT SEQUENCE, 3
            unordered pieces.
ACCESSION   AC121746
VERSION     AC121746.4  GI:25138127
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 265001)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 265001)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 265001)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 20, 2002 this sequence version replaced gi:23267584.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GYXS
                Center clone name: CH230-109L10
            -------------- Summary Statistics
```

```
            Assembly program: Phrap; version 0.990329
            Consensus quality: 245039 bases at least Q40
            Consensus quality: 248664 bases at least Q30
            Consensus quality: 250862 bases at least Q20
            Estimated insert size: 243295; sum-of-contigs estimation
            Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
          --------------
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 3 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *       1     3542: contig of 3542 bp in length
          *    3543     3642: gap of unknown length
          *    3643   263657: contig of 260015 bp in length
          *  263658   263757: gap of unknown length
          *  263758   265001: contig of 1244 bp in length.
FEATURES             Location/Qualifiers
     source          1. .265001
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-109L10"
     misc_feature    1. .1128
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    complement(2758. .3296)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:
                     end_sequence:BH301116"
     gap             3543. .3642
                     /estimated_length=unknown
     misc_feature    3643. .4673
                     /note="wgs_contig"
     misc_feature    253888. .254679
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:BH301114"
     misc_feature    262594. .263657
                     /note="wgs_end_extension
                     clone_end:T7"
     gap             263658. .263757
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              37.5          Length:         265001
Score:                  77.00         Matches:        11
Percent Similarity:     100.00%       Conservative:   1
Best Local Similarity:  91.67%        Mismatches:     0
Query Match:            91.67%        Indels:         0
DB:                     14            Gaps:           0

US-09-858-332G-15 (1-12) x AC121746 (1-265001)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              :::||||||||||||||||||||||||||||||||||
Db      68111 AACAACCACAACCACAACCACAACCACAACCACAAC 68146

RESULT 197
AC118765/c
LOCUS       AC118765              287652 bp    DNA     linear   HTG 23-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-86A14, *** SEQUENCING IN PROGRESS
            ***, 4 unordered pieces.
ACCESSION   AC118765
VERSION     AC118765.4  GI:25188445
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 287652)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 287652)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-APR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 287652)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 23, 2002 this sequence version replaced gi:23321929.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
```

```
          (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
          in the feature table below represents a scaffold in the Atlas
          assembly (a 'contig-scaffold'). Within each contig-scaffold,
          individual sequence contigs are ordered and oriented, and separated
          by sized gaps filled with Ns to the estimated size. The sequence
          may extend beyond the ends of the clone and there may be sequence
          contigs within a contig-scaffold that consist entirely of whole
          genome shotgun sequence reads. Both end sequences and whole genome
          shotgun sequence only contigs will be indicated in the feature
          table.
          -------------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
          -------------- Project Information
              Center project name: GUGT
              Center clone name: CH230-86A14
          -------------- Summary Statistics
              Assembly program: Phrap; version 0.990329
              Consensus quality: 235491 bases at least Q40
              Consensus quality: 241286 bases at least Q30
              Consensus quality: 244445 bases at least Q20
              Estimated insert size: 235593; sum-of-contigs estimation
              Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
          --------------
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 4 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *       1     5124: contig of 5124 bp in length
          *    5125     5224: gap of unknown length
          *    5225   255858: contig of 250634 bp in length
          *  255859   255958: gap of unknown length
          *  255959   286473: contig of 30515 bp in length
          *  286474   286573: gap of unknown length
          *  286574   287652: contig of 1079 bp in length.
FEATURES             Location/Qualifiers
     source          1. .287652
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-86A14"
     misc_feature    1892. .2900
                     /note="wgs_contig"
     misc_feature    3790. .5124
                     /note="wgs_contig"
     gap             5125. .5224
                     /estimated_length=unknown
     misc_feature    5225. .6269
                     /note="wgs_contig"
     misc_feature    126513. .127616
                     /note="wgs_contig"
     misc_feature    131095. .132687
                     /note="wgs_contig"
     misc_feature    172566. .174021
                     /note="wgs_contig"
     misc_feature    183886. .185326
                     /note="wgs_contig"
     gap             255859. .255958
                     /estimated_length=unknown
     misc_feature    255959. .257363
                     /note="wgs_contig"
     misc_feature    270889. .272025
                     /note="wgs_contig"
     misc_feature    277116. .279120
                     /note="wgs_contig"
     gap             286474. .286573
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              40.6            Length:         287652
Score:                  77.00           Matches:        11
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  91.67%          Mismatches:     0
Query Match:            91.67%          Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-15 (1-12) x AC118765 (1-287652)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              :::||||||||||||||||||||||||||||||||||
Db      42251 AACAACCACAACCACAACCACAACCACAACCACAAC 42216

RESULT 198
AC116979
LOCUS       AC116979               331039 bp    DNA     linear   INV 21-FEB-2004
DEFINITION  Dictyostelium discoideum chromosome 2 map 6445720-6776760 strain
            AX4, complete sequence.
ACCESSION   AC116979
VERSION     AC116979.3  GI:42733765
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 331039)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 331039)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 331039)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 331039)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   5  (bases 1 to 331039)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2004) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On Feb 21, 2004 this sequence version replaced gi:28829703.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES             Location/Qualifiers
     source          1. .331039
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
```

/chromosome="2"
/map="6445720-6776760"
CDS complement(join(912. .1165,1280. .1427))
/note="GeneID exon scores (in order of location ranges): 30.46, 12.19 - GSCJ_ID dd_02783"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAS38697.1"
/db_xref="GI:42733766"
/translation="MSEKTFTRQEVAKHCSLNSLWIIYNDDVFDVTNFVVEHPGGEEV LKGNGGKDATQEFDDVGHSASAIAKMQSLRIGRIEGAKPREEKKKEIKKTTTTTSAPK QQESAGLGLLKIPLIIIVLAIAAYFFMGDQQ"
CDS complement(join(2345. .3098,3189. .3526))
/note="GeneID exon scores (in order of location ranges): 98.40, 31.82 - GSCJ_ID dd_02784"
/codon_start=1
/product="similar to Dictyostelium discoideum (Slime mold). Protein disulfide isomerase precursor"
/protein_id="AAO52220.2"
/db_xref="GI:42733767"
/translation="MKILLFVTLIALAFVALCSAEGNVVVLSPDNFDTVVDGSKTVFV KFYAPWCGHCKKLAPDFEILADTFAPVSNKVVIAKVDCDQADNKALCSKYDVSGYPTL KIFDKSTTAKDYNGARSVDELLTYINNHAKTNVKVKKAPSNVVDLSPSNFDSVVLDKS KNVLVEFYAPWCGHCKKLMPDYEILGNTYANEKDVVIAKIDCDAADNKAICSKYGVTG FPTLKWFGKQSKDGEKYEQGRDLDTFINYINKQAGVNRVKGGKLAVGAGRVEQLDTIA TEFIAAAAEVRKELVKKAQTVVDSLPEELRTEGSYYVKVMKTIAEKSIDFVTTEIARI TKLVSGSMSGKKADEFAKKLNILESFKSK"
CDS join(5795. .6179,6292. .6567,6659. .6881,7332. .7363, 7523. .8215,8284. .10451)
/note="GeneID exon scores (in order of location ranges): 14.11, 5.65, 10.80, 0.73, 45.28, 141.12 - GSCJ_ID dd_02785"
/codon_start=1
/product="similar to Dictyostelium discoideum (Slime mold). Hypothetical 127.0 kDa protein"
/protein_id="AAO52221.2"
/db_xref="GI:42733768"
/translation="MNINLLKFFIIYYFCFVLINFCNCQTLKITDISNPKNYKVYPQY SDREFCDFSFEILCYVEPAMDEGDQLMFTTDFDYSEMIRSDRTGSIIGITHTALSFGS YTSNIVCYLLNNPTIKVDLNISYSCEMIDYELMKVKILKPNYFKKNNLYGNGIIQLIG LKYPLNNLVIFGGSAYPLNPNYFLISLEELDGGTVENYNISLLFYGPHTIVVTIPAST YLDYNLDISNSKINYYPNETTNVSEFGLGCSTVYSVTINKTSNDYEPYLYVDDGNSKN ADFTPIYQDENFITYLGQFNSFVTVTSPYAQILNSSILSTLLEFECKLLFDESYLFRF KVNTNGYGASIYLGSDITLRYESVVDSESNSSVLEFETVVDLYKDSITKFTISNLLSN RINYLRGEYYSISPLLKFEIPFGELNPVLITNISFLYNNIDVTNKSVNNIIYFSYEGE FDKGSSFALLLNDPVSILKNGDLDSSDFRFSKWNSSTSMFQIEFTVPANTQPGHIPFK LYGNSNPNYNPIDSTMLPLSAQLFIISSNFDSYGPIFNKVEKINTTNEFGWKFVIGDK INGFDYGDIIVRGEMDSSTYKFHLTPLNLTSGDKYNGEYQINIKISILCASQNYIITD VYLFDTRGNFAKFSVSEQCESIRNPFINFLGDSTINKIYKKCTTENDGYDISPPELTL FSAVKTILNDSQTIVIAFDFQAVDLDTGLKNEQYPIVYIESTELQIIQCESKIKSINS KTANYTCEIELPVGFGYNSDIVFSVYGFINNGGYYSGFPSTYISSAFPIYSMSNIQLV RKIEILRTSQISSTGGELWIVGRGFKTSIQYVQIKYSTDLVFSQISFPSIIFSTALLI LDIDPTDKPFTIRVVTKNGQQSNEFIVTPLFSFLSVSPLNTPTPTTTTQLPTPTNKPQ TCLGEPICGGPKQGFCSSKGCVCYQPWIGNDCNAQVVIIPQPSTNTSQPSTEIPIIDN NNNQQPSGINYLYKSLISIVSIRELDFNGNQINLYPFDRWIFTQINQNKSQYFTSIKN NSTTPFNTNITVILEWFKDTSTIEFANSSITMYPSSIKYTIEITEYKFSNQLNNLQLV MSASFESSKTKEICSLKDFGETSNGDNSNYFKIQINDHSLYGRFIKRAIIDSKISTIE NVLLDSKMNSIQTASSSQSYIGITIPNYQKSIIIDPDFSVLIDSKSTSKNDENSVCTP NQSKLTTPQLIGIIIGSIAFAAVIAISIAYHIIKKKKDTKFKNGLQNKLKTFKS"
CDS join(11030. .11042,11114. .11283)
/note="GeneID exon scores (in order of location ranges): -1.76, 12.26 - GSCJ_ID dd_02786"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAO52222.2"
/db_xref="GI:42733769"
/translation="MTILKSISSMGISNRKNSFSFVSTNSMIETNQNVNSVSKSGYSK LIKGAALMAEGISGFF"
CDS complement(11515. .12534)
/note="GeneID exon scores (in order of location ranges): 87.29 - GSCJ_ID dd_02787"
/codon_start=1
/product="similar to Burkholderia sp. DBT1. Putative oxidoreductase"
/protein_id="AAM34303.3"
/db_xref="GI:42733770"
/translation="MLPITNKVALITKKGFSESPSIVDNEIPKPKGNEVLIKVDYVGL NRCDLYFSLGLYLPFENYPNVSGFEASGTVVEFANDELENSSKFKVGDKVSVIYTCDH RKYGTFAEYAVFPIGCLVKNNELSSQKEIAANWISFFTAYYGLVEFAQIKKGDYVIIT AGSAGTSLAGIAIAKFFGAHTIVTSRNEKKKQQILDYGADYFVSLDTEDIVKRTFEIT NGKGANIIYDSVGGQSAQKLFESSSQFGKIVFYGNMNHSEPTPLPILIGLKNYLTIKV LNTNEYLYQPEFNKKAIEFLNSNFHHFKSIVGKEFIGIDSLVDALKYLETGDLFGKVV VKVNN"
CDS complement(join(13231. .14884,15364. .15368))
/note="GeneID exon scores (in order of location ranges): 104.49, -7.13 - GSCJ_ID dd_02788"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAM34304.3"
/db_xref="GI:42733771"
/translation="MINPNLLKPLFIENFCSSSQSWLDLIKNRKLNFYTITCLTHYYS TVKVLYFDIKSIFENDKKQNENQQNEKQQQQQQQQQQQQQQNEKSLNDRKSIFKKSYH IIRLLIEMNDIINYKDIRIYLPNQSEERKLLMKIREGEEIDIQSCFELVQSKYLEFES KFNELKKNSTNDSLIKLNSARALHVLNDWVVDIRRNSIENSNNLNSFKKYKFENENHH DESNNSFKIAMDLMNKNNIKGTLIYFGPSGSHLHGLQDDDHNDENSNDNEKIDWIGIY AAPTDDIVSLFPLALKHDSSGFIMQSNEHIGINLNPAPDRPNNQNKKDSFSKGIILYE LKYALQLIVNCSYRVLECVFMSSNQVLNNHPFKNKFYITDAWKDFINQVLINDNNNNK TFSKLSNSNLIQQYWGLAKGISMKNSTIFKMVKSNLLRKKDDAITIAEEDLVTDDSIS KNLFFLNYLLIQCNTILNGIVFNENNTIINSPNNEILLKEKEIPSIFIDKQSKDYQFL MSLKHNKLNKLSIIQSLIIDSDSILEEIRIKLDTLKKIQRIDLV"
CDS join(17515. .17664,18062. .19243)
/note="GeneID exon scores (in order of location ranges): 1.18, 93.30 - GSCJ_ID dd_00463"
/codon_start=1
/product="similar to mitochondrial carrier protein family; protein id: At1g14560.1 [Arabidopsis thaliana]"
/protein_id="AAO52223.2"
/db_xref="GI:42733772"
/translation="MNSSDFKKSFKESTENNSNTYRPSKTLNTYKSLKNGNNIANSLI IQLNNNYLVSGSIAGAISRSATAGFERLTIIQQVQGMSQNLSQGYVGCIAAMKEMVKR EGFKSIWKGNGANIVKVSPNSGIRFLTYEFCKKHFLDNSSNHPSSSSIENGIDGNGVG CGSGSEMKMTVPQTMFSGAMAGLTSTFFTYPLDVVRIRLSLQGSCSNDYAAHRYNGIT HSFFKIHKDEGVKGLYKGLGTSIASIVPWVSISFATYEGFKIICKKMILNYQISSSSL STTTTPSSINNNNNNNNKNNNSFIYENELGENGINLTNTSGCSTMASTMPSSLLINS VASDENELKKGVNMICDFVCGALSGAVTMTVCYPLDVLRRRMMIQGIGGNKVLYKNGW DATKKILSNEGLVAFYHGIIPAYFKVVPTVAISFAVYEICKDLGSNKYQQK"
CDS join(20299. .20406,20551. .20660,20960. .21569,21708. .21828, 22984. .24188)
/note="GeneID exon scores (in order of location ranges): 9.06, -2.15, 55.50, 6.70, 58.02 - GSCJ_ID dd_01980"
/pseudo
/codon_start=1
CDS join(24256. .24266,24388. .24538,24623. .24750,24902. .27884)
/note="GeneID exon scores (in order of location ranges): -2.62, 6.90, 11.20, 259.32 - GSCJ_ID dd_01978"
/codon_start=1
/product="similar to Dictyostelium discoideum (Slime mold). Histidine kinase DhkF"
/protein_id="AAO52224.2"
/db_xref="GI:42733773"
/translation="MILKFIIQQCVFKVAKIFSDNLIQIFMDISSVYSSSTLFFSKLS LLMCKVFNSKCAGISVLNKDGLFDTIAFCSSTQELENFSYSLKKISKNITYYDYMEIP LKDSKEFYFGDIVLFGDQTLNPNILDTNILTFITNRTSMEFEKKKVSDQLIIAKNLLD QSPSCSCLVSRNGKIIRKIGAFYERLLGLEIGENISSVEKSSDISTILKQVCFENNNL STETTITKKNGEKYPAEVFVKEISDIHSNSIGIMIIVRDITDQIRLKEMNIELQKKSE LEQKRNQELMEARDLALTATKIKSQFLATISHEIRTPLNGIITMGEMLLSTSPLNTEQ HDIAETIFGSSELLLSITSDILDFSKIEASKLELEMIEFDFIGCLEGIGKTIGVSITN KPIEIAFLMDTDIPHRLIGDPNRLIQIMLNMGTNAVKYTDRGHIVFRISVISREQNRC KIKISIDDSGIGIGEDQRAHLFEPFHQIDSSSTRKYGGSGLGLAISSKLAKLMGGEVI LERSKPGVGSLFSVTLNFEQIGSNTLKSLLPDKSFSRNKTCIILDNYEYTASIASQRF DQIFEESNIILVKESTVKKFFELVQEYKINGDVNKIDLKNLDPELLPFFDSSLICIIV FHRFMDNLDIFLKYVKDFIEFYKKKIVVALGINHKNFKNLPKRRDFLIFKKPISSTNL IKVSNMARKIIPKSSSSSNLIQTISQIDNQQQQQQQQLQQQEQEQQHQQQQLQQEQQF VTISPHSDSSEKKTTPKKDRGKYDFNISPLRIDRFGTESTSSPKIKLYSDTSSDSGES DEFEFSENLRELKSDGELVVLKNQPYKEQRLSNIQLINDPIVTPPTPKANSEVVSTKA ESTFFQSSSITIKETRSDSVNSVNSVNSVNSVNSLNSENSNNSLDVGIRTRPRILLVD

```
                    DNAVNRKVVKLQLKKLGYDCDTATNGFEGFEMQKKDNYELIFMDLNMPLCDGSMASKL

Alignment Scores:
Pred. No.:              46.5          Length:         331039
Score:                  77.00         Matches:        11
Percent Similarity:     100.00%       Conservative:   1
Best Local Similarity:  91.67%        Mismatches:     0
Query Match:            91.67%        Indels:         0
DB:                     2             Gaps:           0

US-09-858-332G-15 (1-12) x AC116979 (1-331039)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||:::|||||||||||||||||||||||||||
Db     317422 CACAACAACAACCACAACCACAACCACAACCACAAC 317457

RESULT 199
AC116986
LOCUS       AC116986              333321 bp    DNA     linear   INV 12-MAR-2003
DEFINITION  Dictyostelium discoideum chromosome 2 map 2234041-2567370 strain
            AX4, complete sequence.
ACCESSION   AC116986 AC116425
VERSION     AC116986.2  GI:28829247
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 333321)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 333321)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 333321)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 333321)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On or before Mar 4, 2003 this sequence version replaced
            gi:19774405, gi:20042949.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES            Location/Qualifiers
     source         1. .333321
                    /organism="Dictyostelium discoideum"
                    /mol_type="genomic DNA"
                    /strain="AX4"
                    /db_xref="taxon:44689"
                    /chromosome="2"
                    /map="2234041-2567370"
     CDS            complement(join(2281. .4838,5022. .5160))
                    /note="GeneID exon scores (in order of location ranges):
                    218.28, 7.77 - GSCJ_ID dd_03279"
                    /codon_start=1
                    /product="similar to Plasmodium falciparum (isolate 3D7).
                    Hypothetical protein"
                    /protein_id="AAM08764.2"
                    /db_xref="GI:28829248"
                    /translation="MKKLFDIISPSKNIQQQSASGKESSSSSSSSSSPSPSKNRSCFN
                    GIGINSKINNINIQEDSISNGNNNEKYPQIERDNIIKIVLLGGYQVGKSSIIRRYLQG
                    CFTTSYIPTIGIDLSTLLLKRKQSQQQQQQQQQQQQQQQQQQQQQQQQQQQPSPQSSQQPPPQ
                    QQQQQQQQQQQQQQQQTPQRNKNENYDENDYLSDYSDSESKDDSSLNQYLTQSPNQYIV
                    QFWDVAYAEIKGKHFSSILSGASAIIMVYDSVNASSILAIDEWKTILNTSPFIQINNG
                    NNNNNNNNNNNNSNNSNNNSNNNSFNSNNYKKIPLALFANKSDIGEPVISPTDLD
                    YYCNACGYSIWRFTSCKKDLNIISGIDSLISLVIENKNQELSDIRNRLLNEKKEKRKL
                    IYQLKKNVNLNTSVKYTRSINKSSSGNKNQDNNNVDDIDIVKNLTKELVLSQSCPNIS
                    NYGKDVNNDDDDDNNNTIENNENENNENVKKQQEKEKEQKEKEKEKLQQFKHQSEIEK
                    VNELLEMTNEHYSKFQKSLEEIMSKSNGTTRYNDLVQLEKQRSNEYVKICNNLKKLIQ
                    PSAFRQHHHHHHHHHHHKPSQKITSDLLSKILDENILKWNTLISNLQLEISLKLNNQN
                    NVYQSNNNNNNNVKQNNKVAQSYNAVSLLIKDDQEYRYNNNNNNNNGGRKNSFSNSS
                    NNINNNNNNINNNNNNNNNNNNNNNNNNNNNNNNNNSSINNSIKLNINNNPKYKMG
                    FGFKHGGKNLSSNFSINDHSIDTDNPSMTSSSITSNSSMNSFSSSPSSSPPNDDINDN
                    RHTNNINRVATNGDGNVDGGGGDNIGDNDSSSLNSSSYQNHHNHNHNHRHNFNNDSNN
                    NSYSPSTSHNGIGGDDFLLPSPSLLPSSYNLTVTVLPDLKYK"
     CDS            join(5597. .5760,5890. .6593,6692. .10869)
                    /note="GeneID exon scores (in order of location ranges):
                    7.48, 52.43, 406.35 - GSCJ_ID dd_01879"
                    /codon_start=1
                    /product="similar to Arabidopsis thaliana (Mouse-ear
                    cress). Putative UDP-glucose:glycoprotein
                    glucosyltransferase, 101200- 91134"
                    /protein_id="AAM08766.2"
                    /db_xref="GI:28829249"
                    /translation="MARIFKFFVFLLIVFISNVLLLVESNEGDNSFSSKSIQLSLVSN
                    WGETPSYLEAAEFLHNQDKSLFWKFIEEFNKIDFSTNYSDKIYYESTISLMKSVLSSN
                    TQFLSEFLSIDLAMRTYSPRVETYRQLAISNMKLNNIEHSITTADNKTITLFNSGGWV
                    QIKNKIITDVNEINESLFKDVAVVDDEENEFIRLYDFDHIFPTLANTVSSSSSSPSSI
                    PIVILYVDIKSEFFKLVHPKLKQFSQMGKIKYCLRYVVQESNQKLNLQGYGYELSIKN
                    LEYKVMDDSAIKKDIIIDGVKSKTIINIPNEDVQGFNFHKLQKRKPELTSKLSTFRSY
                    LMAKSQEAKELKVWELKDLGIQSAQKIIQSGDPLRSLEYISQKFPTLSNSLSKITLNE
                    SLKSVIESNQKIIPSTTDQTLLLNGRLIDTNELSPIELSRIILEEYEHSTTIQQQGPL
                    SSKTVQDIISAQLPIRIQLLPTKEELELNGGNEPFVSLNNLELDYIYRQWEPKLQSSV
                    LDKPVTSPQDIFIRKNLLTTVIVLDWNNINTFEIIPEIQEMVQGNSLIPTRIQLLFNT
                    KSNNNNNNNNNNNDQNSQTSNFIQGKDLAKVFLTIKNSNLGNRGAFFFITALNYFKKM
                    YIPNELGITRSVLSSSFQAVLQQMGGSVRSLQHALTNTDFDNLLESSNQLIERLELLD
                    TTTSQSTTTTTTTKILPKVFVNGVQVKYSNIDQLSFDLLVSLYDEFDNLKPLFKESIL
                    STTTAQYYETILTSSYWKDNNLPFLKKLNSMISNEKYSHLITNSKNRNQEVDAQNVLK
                    NLLYFRNNENKDEQNLLNLIVIGDFDHYNTRDISLELLRQLEKGELKNCKLTFISNPI
                    DINSVVNTAGNENQILGKLITILKHYGKILTPQLVIGLFEKVQSDPTIIDSFKTMKQI
                    IELSGFDIAANDIWVAQSVNLFKQSSKVCKQYLGIQSTNKSPLSILVNGRIITPPLSY
                    DDAASFIQSDFKLLLEIEMIKAKKTFELLNSDPILKDKSNLKISDLLNKVQSLVGYYY
                    NGNNQLDSNIKRKRIPNSLSISFSHKPPTLSSSSSSSSSNSNDVPLKFLMIINPFNKV
                    SQKLVPMVREFSNKLNIPVDVILNPPVSLSELPLKTYYTYVIKLSSEFNNENVLYNQP
                    LGIATDIPEDRVVTLALDIPSSWLVQPIIAKYDLDNIRLKDLGDEQVLTAVYELENIV
                    IEGSANDMTTDNAPAGLELLLNPISTQTNKTQDTIVMNNFGYYQLKSNPGIWKLTIAP
                    GRSSDIMDMVDHPNQKEKETFVIVPHRLVVIDSLYQSLSSLSVVRKAGQELRPILQPI
                    DEYEKQKEQEKEQKLKQNSSGFFSNLFSSKNDATDSVATHQKKSNLDTIHIFSVASGH
                    LYERFLKIMMLSVVKNTESPIKFWFLKNYLSPAPKEFIPEMAKEYGFQYELVTYKWPW
                    WLRKQTEKQRIIWSYKILFLDVLFPLDVPKIIFVDADQVVRTDLKELWDMDLHGASLG
                    YTPFCDSNKDTEGFRFWKSGYWRQHLAGRSYHISALYVVDLVRFRRLAAGDQLRATYD
                    QLSRDPNSLANLDQDLPNYLQHYVRIHSLPQEWLWCETWCDQESKSKAKTIDLCNNPL
                    TKTPKLENAVRIIDEWTTLDNEAKEFELKIDQSKHHRQIELDHQNQLPNSKPIENIDD
                    ILLNLAESQKDLF"
     CDS            complement(join(10993. .12657,12738. .12853,13014. .13262,
                    13781. .13814,14302. .14529))
                    /note="GeneID exon scores (in order of location ranges):
                    175.98, 4.12, 21.10, -3.61, 20.21 - GSCJ_ID dd_01880"
                    /codon_start=1
                    /product="similar to Neurospora crassa. Probable branching
                    enzyme (be1)"
                    /protein_id="AAM08768.2"
                    /db_xref="GI:28829250"
                    /translation="MNVDDIYADIIIESENDQENDMKQSDENSQSPTLKEDHHNKIKE
                    FEEIIENNLKEISNLKKEIESKDNEIETLSKQIFSIIVKLILRDYSKATDGTKVIHDD
                    PWLEPYKEVIKRRHNQVKNTIQKLEESEGSLLKFSQGYEYFGFNVTKDGVNYREWLPS
                    AHEVYLVGDFNQWNKTSHPLERDNYGRWSIFIPNNSNGECAIPHGSKIKIYLKLANGN
                    FDYRIPAWIKRVEQTKENPVFDGVFWNPSKQYVFKNKSPMKPTELRIYEAHVGMSSEL
```

```
PEISTYSKFKDTVLPMVKELGYNCIQLMAVMEHAYYASFGYQVTNFFAISSRFGTPEE
LKEMIDKAHEMGLLVFLDVVHSHASKNVLDGLNQLDGTDHHYFHSGGRGNHELWDSRL
FNYGNWEVMRFLLSNLRFYVDEYHFDGFRFDGVTSMIYTHHGLSPACSYDDYFGGAVD
EDALNYLTLANVMLHTLNPSIVTIAEEVTGLATLCRPFSEGGGDFDYRLAMGIPDKWI
ELVKEKKDEDWNMGTIAHMLSNRRYKEKNIAYAESHDQSLVGDKTLAFWLMDKEMYTN
MSVTTEETPIIDRGMSLHKMIRLITSSLGGDGYLNFMGNEFGHPEWVDFPREGNNNSL
HHARRRWDLYRNPLLRYKQLRDFDIAMNKAEQEFRWLSSDFAYISLKHEDDKIIVFER
ASLIFIFNFHPSKSFSDYRIGSGVPGKFINVLDSDRKEFGGHVRIGKDNYHYTEDKPW
HDRKYSLLIYIPSRTCLVLKKVD"
CDS     join(15053. .15120,15221. .15743)
        /note="GeneID exon scores (in order of location ranges):
        -0.56, 37.71 - GSCJ_ID dd_01882"
        /codon_start=1
        /product="hypothetical protein"
        /protein_id="AAM08769.2"
        /db_xref="GI:28829251"
        /translation="MGQKSLGELQYPENMKPLFNVKIEDCKYSYDTKDDTFLSIQYND
        NNKQAKSNNDSSIGNNKNSKEQSSSTTTTTRRRNVNSKEPIIKTSSNGVTTMGSLESL
        LNNCSTNDNSSDDDDDDESDTDEEDEDNDNSFENNPNLTEEEKQKIFQEKVKKQQQKK
        KLKANKKIIRKDPIYWFGISTPTSLKQSQSYFKQGK"
CDS     complement(join(16609. .16729,16860. .16934,17034. .17080,
        17169. .17177))
        /note="GeneID exon scores (in order of location ranges):
        6.96, 13.18, -2.05, 0.27 - GSCJ_ID dd_01883"
        /codon_start=1
        /product="hypothetical protein"
        /protein_id="AAO51794.1"
        /db_xref="GI:28829252"
        /translation="MAGPWTIEPGKNHGGIPTWAWPRSTTVHVQVVGGGRGRVRMQAG
        ASPDEDNDVNGETSFSRSFGGSRLNVTNIGSRTLKVWTA"
CDS     complement(19502. .21007)
        /note="GeneID exon scores (in order of location ranges):
        108.08 - GSCJ_ID dd_01875"
        /codon_start=1
        /product="similar to Dictyostelium discoideum (Slime
        mold). FutA (Fragment)"
        /protein_id="AAO51795.1"
        /db_xref="GI:28829253"
        /translation="MKSLRNSYSILLKIVIISTCLLLLINIALIYEDPNLIDSNCKTT
        TTITKTQNNKNNKGLENLNKSKKKLLLILDYKLFNEDTYMKSFSECTSITNQPQCGFT
        TDISLANESDGILYFIPRTTKKNYNRSYTSVHRTSEKQLFFGWSMESPSMYPLENDDT
        HLSLNYNVTIGYPLNDYQKYGAGVSDEEYEEKDIQSSHIYTPYGPVPFKGSDHYSYLF
        KFGKLKEIPKKRDSNIMWMAGNCRNQKLGRTKLVKQMMRFTTIDSYGKCLHNKDLKVN
        YGPNVLLRSLQKLKILRKYDFVVAIENSRCEDYVTEKLWEPLSVGTIPIYLGATNIDE
        FLPHPDAIINIANFNSINHLMKYVRDVGLNETLREKHLAWIKKPLPKKFKNLLNQSLN
        NQNLFCSICKKLVNLDNNDNINYTPLSKPFKQCIEGKFTLPIYTEEEKERGRLLRLQK
        QQEQEKLELQNNNDNYTNDNEEEEEEEDNDSGNDSGNDSGNDSGNNDSNSI"
CDS     complement(join(21960. .26819,26946. .27251))
        /note="GeneID exon scores (in order of location ranges):
        530.44, 21.78 - GSCJ_ID dd_03196"
        /codon_start=1
        /product="similar to Dictyostelium discoideum (Slime
        mold). Nucleotide exchange factor RasGEF R"
        /protein_id="AAO51796.1"
        /db_xref="GI:28829254"
        /translation="MKEEEIANLGPQEYFNYLEGINKKIIIELKTVANIDTAKDTDNL
        QNILIESSLALEHNIRNLLYFVDKCPISYDQIVHYHSHFKEVVKIFFQVCQDLIKAAQ
        ALILNSCDYLTTSNFAKCRGEAIKTIKNTMATCSNFEKLYYNPDLQLEDEVDLVHQQQ
        QKEQEQQQKEQEQYEQEQLILKQLKEQELKKQQEQQQQQQQQQQQQQQQQQQQQQQKHQ
        EEKEKNDQKEKEEKLKKQLLAEQQKQKQEELERQQQQLQQQQQLLQQQQQLLQQQQQQ
        QRSDPIIIPPITLKTSQSDLLSGGLSPKVNSPGSPNLGKFTESAERVANFDRGTANII
        LNMNKIKEISLSGDSEQAPEIVTAARIISENIAIISKELRYKTLGTNLSEHFVSVVQV
        ARLAIKNKSDQFYQDQLELAIEKFNDTMRKIIYSVKSISKSSMNLRQFGVDESSLGRS
        PVVSPEKSISPSFTSSTSERILSHEIKNHNNNNNNYNNSSTNNLQTSFSTPSLSSNHS
        QQPNQQPLQSPLLINQLQSTSSSSSSSSSNLSNSLNSIQLPQATTTTTTATSPSTSTS
        TSTSTSPNSSSLSISDQDKQLKRKEKQHQIVIPKDVKDSVANQQQQQQQQQNGTTSPR
        NNESSVTAATTTTTSTTASITTNVNTIPNFPPNKQLPATPTTGTPSTSTPTPQTPTST
        SQNDKQNENNNKENFVDKQKTLGKLFSKFVHKKRTPLPGFDSSNSSSPSNNSNTTNSS
        SHSSTNSSPMETSPGVESPKITKSPSQNNILVDSLDIGSDNQQQQQDKLTTTTSTTTI
        TINNNNNNNNNNNNNNNNNNNIQQQQQQQQQIPTTPNKLSNSIAMNSSNRLMTPKKE
        RSFTIGLVSGKHSCVVVETPKSKKLFQHESAKQILSIISMNFPSFEKEFQMQNSEVIS
        NIIEQIGNVIKNYHDEVSESSSSSSSSTTSPNNINTPSDCSPILNSENSKDSHNLSSI
```

```
Alignment Scores:
Pred. No.:                46.8          Length:          333321
Score:                    77.00         Matches:         11
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    91.67%        Mismatches:      0
Query Match:              91.67%        Indels:          0
DB:                       2             Gaps:            0

US-09-858-332G-15 (1-12) x AC116986 (1-333321)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||:::||||||||||
Db     209485 CACAATCACAACCACAACCACAACAACAACCACAAC 209520

RESULT 200
AJ811567
LOCUS       AJ811567                 122 bp    mRNA    linear   MAM 13-OCT-2004
DEFINITION  Lama glama partial mRNA for immunoglobulin heavy chain hinge region
            (ighc gene), clone pRL188.
ACCESSION   AJ811567
VERSION     AJ811567.1  GI:51490834
KEYWORDS    ighc gene; immunoglobulin heavy chain hinge region.
SOURCE      Lama glama (llama)
  ORGANISM  Lama glama
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Tylopoda;
            Camelidae; Lama.
REFERENCE   1
  AUTHORS   Harmsen,M.M., Van Solt,C.B. and Van Keulen,L.
  TITLE     Selection of llama single-domain antibody fragments binding to
            recombinant bovine prion protein in ELISA
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 122)
  AUTHORS   Harmsen,M.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-AUG-2004) Harmsen M.M., Infection Biology, Animal
            Sciences Group (ID-Lelystad BV), Edelhertweg 15, Lelystad, 8219 PH,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .122
                     /organism="Lama glama"
                     /mol_type="mRNA"
                     /db_xref="taxon:9844"
                     /clone="pRL188"
                     /cell_type="B lymphocyte"
                     /rearranged
     gene            1. .122
                     /gene="ighc"
     primer_bind     <1. .25
                     /gene="ighc"
                     /note="primer LH derived"
     CDS             <17. .>93
                     /gene="ighc"
                     /codon_start=1
                     /product="immunoglobulin heavy chain hinge region"
                     /protein_id="CAH18863.1"
                     /db_xref="GI:51490835"
                     /translation="EPKTPKPQPQPQPQPQPNPTTESKCP"
     primer_bind     71. .>122
                     /gene="ighc"
                     /note="primer BOLI269 derived"
ORIGIN

Alignment Scores:
Pred. No.:                0.0309        Length:          122
Score:                    76.00         Matches:         11
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              90.48%        Indels:          0
DB:                       4             Gaps:            0

US-09-858-332G-15 (1-12) x AJ811567 (1-122)
```

```
Qy          2 AsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||
Db         33 AACCACAACCACAACCACAACCACAACCACAAC 65

RESULT 201
AF097205/c
LOCUS       AF097205                390 bp    mRNA    linear   MAM 17-DEC-2003
DEFINITION  Bos taurus clone BF2B5 immunoglobulin heavy chain variable region
            mRNA, partial cds.
ACCESSION   AF097205
VERSION     AF097205.1  GI:3834650
KEYWORDS    .
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 390)
  AUTHORS   Saini,S.S. and Kaushik,A.
  TITLE     Extensive CDR3H length heterogeneity exists in bovine foetal VDJ
            rearrangements
  JOURNAL   Scand. J. Immunol. 55 (2), 140-148 (2002)
   PUBMED   11896930
REFERENCE   2  (bases 1 to 390)
  AUTHORS   Saini,S.S., Allore,B., Jacobs,R. and Kaushik,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1998) Pathobiology, University of Guelph, Guelph,
            ON N1G 2W1, Canada
FEATURES             Location/Qualifiers
     source          1. .390
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /clone="BF2B5"
                     /cell_type="mouse x Bovine heterohybridoma"
                     /dev_stage="fetal"
                     /note="PCR product"
     CDS             <1. .>390
                     /note="VDJ region"
                     /codon_start=1
                     /product="immunoglobulin heavy chain variable region"
                     /protein_id="AAC71040.1"
                     /db_xref="GI:3834651"
                     /translation="QVQLRESGPSLVKPSQTLSLTCTVFGFSLSDNSVGWVRQVPGKA
                     LEWLGVIYSGGSTGYNPALKSRLSITKDNSKSQVSLSLSSVTTEDTATYYCARDCYSD
                     YGNGYGCGYGYGSYVDAWGQGLLVTVSS"
ORIGIN

Alignment Scores:
Pred. No.:              0.0947          Length:         390
Score:                  76.00           Matches:        11
Percent Similarity:     91.67%          Conservative:   0
Best Local Similarity:  91.67%          Mismatches:     1
Query Match:            90.48%          Indels:         0
DB:                     4               Gaps:           0

US-09-858-332G-15 (1-12) x AF097205 (1-390)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||   ||||||
Db        340 CATAACCATAACCACAACCATAACCATTACCATAAT 305

RESULT 202
A57711
LOCUS       A57711                  537 bp    DNA     linear   PAT 03-MAR-1998
DEFINITION  Sequence 21 from Patent WO9634103.
ACCESSION   A57711
VERSION     A57711.1  GI:3713535
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified sequences.
REFERENCE   1
  AUTHORS   Hamers,R. and Muyldermans,S.
  TITLE     Variable fragments of immunoglobulins - use for therapeutic or
            veterinary purposes
  JOURNAL   Patent: WO 9634103-A 21 31-OCT-1996;
            UNIV BRUXELLES (BE)
COMMENT     Other publication AU 5647896 961118.
FEATURES             Location/Qualifiers
     source          1. .537
                     /organism="unidentified"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32644"
     CDS             1. .537
                     /note="unnamed protein product; Protein sequence is in
                     conflict with the conceptual translation"
                     /codon_start=1
                     /protein_id="CAA03459.1"
                     /db_xref="GI:3713536"
                     /translation="DVQLQASGGGSVQAGGSLRLSCAASGYTIGPYCMGWFRQAPGKE
                     REGVAAINMGGGITYYADSVKGRFTISQDNAKNTVYLLMNSLEPEDTAIYYCAADSTI
                     YASYYECGHGLSTGGYGYDSWGQGTQVTVSSEPKIPQPQPKPQPQPQPQPKPQPKPEP
                     MECGRYPYDVPDYGSGRA"
ORIGIN

Alignment Scores:
Pred. No.:              0.129           Length:         537
Score:                  76.00           Matches:        11
Percent Similarity:     91.67%          Conservative:   0
Best Local Similarity:  91.67%          Mismatches:     1
Query Match:            90.48%          Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-15 (1-12) x A57711 (1-537)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||   ||||||||||||||||||||
Db        413 CACAACCACAACCAAAACCACAACCACAACCACAAC 448

RESULT 203
A57713
LOCUS       A57713                  912 bp    DNA     linear   PAT 03-MAR-1998
DEFINITION  Sequence 23 from Patent WO9634103.
ACCESSION   A57713
VERSION     A57713.1  GI:3713537
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified sequences.
REFERENCE   1
  AUTHORS   Hamers,R. and Muyldermans,S.
  TITLE     Variable fragments of immunoglobulins - use for therapeutic or
            veterinary purposes
  JOURNAL   Patent: WO 9634103-A 23 31-OCT-1996;
            UNIV BRUXELLES (BE)
COMMENT     Other publication AU 5647896 961118.
FEATURES             Location/Qualifiers
     source          1. .912
                     /organism="unidentified"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32644"
     CDS             1. .912
                     /note="unnamed protein product; Protein sequence is in
                     conflict with the conceptual translation"
                     /codon_start=1
                     /protein_id="CAA03460.1"
                     /db_xref="GI:3713538"
                     /translation="DVQLQASGGGSVQAGGSLRLSCAASGYTIGPYCMGWFRQAPGKE
                     REGVAAINMGGGITYYADSVKGRFTISQDNAKNTVYLLMNSLEPEDTAIYYCAADSTI
                     YASYYECGHGLSTGGYGYDSWGQGTQVTVSSEPKIPQPQPKPQPQPQPQPKPQPKPEP
                     MAEVQLQASGGGSVQAGQSLRLSCATSGATSSSNCMGWFRQAPGKEREGVAVIDTGRG
                     NTAYADSVQGRLTISLDNAKNTLYLQMNSLKPEDTAMYYCAADTSTWYRGYCGTNPNY
                     FSYWGQGTQVTVSSGRYDVPDYGSGRA"
ORIGIN
```

```
Alignment Scores:
Pred. No.:                0.215          Length:          912
Score:                    76.00          Matches:         11
Percent Similarity:       91.67%         Conservative:    0
Best Local Similarity:    91.67%         Mismatches:      1
Query Match:              90.48%         Indels:          0
DB:                       6              Gaps:            0

US-09-858-332G-15 (1-12) x A57713 (1-912)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||   ||||||||||||||||||||||
Db        413 CACAACCACAACCAAAACCACAACCACAACCACAAC 448

RESULT 204
AX164109
LOCUS       AX164109                1154 bp    DNA     linear   PAT 31-AUG-2001
DEFINITION  Sequence 54 from Patent EP1118669.
ACCESSION   AX164109
VERSION     AX164109.1  GI:14545059
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Frenken,L.G., van der Logt,C.P., Jobling,S.A. and Teh,Y.M.
  TITLE     Production of camelid antibodies in plants
  JOURNAL   Patent: EP 1118669-A 54 25-JUL-2001;
            UNILEVER PLC (GB) ; UNILEVER N.V. (NL)
FEATURES              Location/Qualifiers
     source           1. .1154
                      /organism="synthetic construct"
                      /mol_type="unassigned DNA"
                      /db_xref="taxon:32630"
                      /note="HCV33-hinge-CH2-CH3"
     CDS              3. .1136
                      /note="unnamed protein product"
                      /codon_start=1
                      /transl_table=11
                      /protein_id="CAC42712.1"
                      /db_xref="GI:14545060"
                      /translation="MEVQLQESGGGLVQAGGSLRLSCAASGRATSGHGHYGMGWFRQV
                      PGKEREFVAAIRWSGKETWYKDSVKGRFTISRDNAKTTVYLQMNSLKPEDTAVYYCAA
                      RPVRVDDISLPVGFDYWGQGTQVTVSSEPKTPKPQPQPQPQPQPNPTTESKCPKCPAP
                      ELLGGPSVFIFPPKPKDVLSISGRPEVTCVVVDVGQEDPEVSFNWYIDGAEVRTANTR
                      PKEEQFNSTYRVVSVLPIQHQDWLTGKEFKCKVNNKALPAPIEKTISKAKGQTREPQV
                      YALAPHREELAKDTVSVTCLVKGFYPPDINVEWQRNGQPESEGTYATTPPQLDNDGTY
                      FLYSKLSVGKNTWQRGETFTCVVMHEALHNHYTQKSITQSSGK"
ORIGIN

Alignment Scores:
Pred. No.:                0.269          Length:          1154
Score:                    76.00          Matches:         11
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              90.48%         Indels:          0
DB:                       6              Gaps:            0

US-09-858-332G-15 (1-12) x AX164109 (1-1154)

Qy          2 AsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||
Db        406 AACCACAACCACAACCACAACCACAACCACAAC 438

RESULT 205
AX164111
LOCUS       AX164111                1172 bp    DNA     linear   PAT 31-AUG-2001
DEFINITION  Sequence 56 from Patent EP1118669.
ACCESSION   AX164111
VERSION     AX164111.1  GI:14545061
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Frenken,L.G., van der Logt,C.P., Jobling,S.A. and Teh,Y.M.
  TITLE     Production of camelid antibodies in plants
  JOURNAL   Patent: EP 1118669-A 56 25-JUL-2001;
            UNILEVER PLC (GB) ; UNILEVER N.V. (NL)
FEATURES              Location/Qualifiers
     source           1. .1172
                      /organism="synthetic construct"
                      /mol_type="unassigned DNA"
                      /db_xref="taxon:32630"
                      /note="hinge-HCV33-CH2-CH3-SEKDEL"
     CDS              3. .1154
                      /note="unnamed protein product"
                      /codon_start=1
                      /transl_table=11
                      /protein_id="CAC42713.1"
                      /db_xref="GI:14545062"
                      /translation="MEVQLQESGGGLVQAGGSLRLSCAASGRATSGHGHYGMGWFRQV
                      PGKEREFVAAIRWSGKETWYKDSVKGRFTISRDNAKTTVYLQMNSLKPEDTAVYYCAA
                      RPVRVDDISLPVGFDYWGQGTQVTVSSEPKTPKPQPQPQPQPQPNPTTESKCPKCPAP
                      ELLGGPSVFIFPPKPKDVLSISGRPEVTCVVVDVGQEDPEVSFNWYIDGAEVRTANTR
                      PKEEQFNSTYRVVSVLPIQHQDWLTGKEFKCKVNNKALPAPIEKTISKAKGQTREPQV
                      YALAPHREELAKDTVSVTCLVKGFYPPDINVEWQRNGQPESEGTYATTPPQLDNDGTY
                      FLYSKLSVGKNTWQRGETFTCVVMHEALHNHYTQKSITQSSGKSEKDEL"
ORIGIN

Alignment Scores:
Pred. No.:                0.273          Length:          1172
Score:                    76.00          Matches:         11
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              90.48%         Indels:          0
DB:                       6              Gaps:            0

US-09-858-332G-15 (1-12) x AX164111 (1-1172)

Qy          2 AsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||
Db        406 AACCACAACCACAACCACAACCACAACCACAAC 438

RESULT 206
AR578923
LOCUS       AR578923                1521 bp    DNA     linear   PAT 14-DEC-2004
DEFINITION  Sequence 7 from patent US 6783985.
ACCESSION   AR578923
VERSION     AR578923.1  GI:56581740
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1521)
  AUTHORS   Roemer,T., Jiang,B., Boone,C. and Bussey,H.
  TITLE     Gene disruption methodologies for drug target discovery
  JOURNAL   Patent: US 6783985-A 7 31-AUG-2004;
            Elitra Pharmaceuticals Inc.; San Diego, CA
FEATURES              Location/Qualifiers
     source           1. .1521
                      /organism="unknown"
                      /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:                0.351          Length:          1521
Score:                    76.00          Matches:         11
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              90.48%         Indels:          0
DB:                       6              Gaps:            0

US-09-858-332G-15 (1-12) x AR578923 (1-1521)

Qy          2 AsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
```

```
                 ||||||||||||||||||||||||||||||||
Db        173 AACCACAACCACAACCACAACCACAACCACAAC 205

RESULT 207
AX240769
LOCUS       AX240769               1521 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 7 from Patent WO0160975.
ACCESSION   AX240769
VERSION     AX240769.1  GI:15797705
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1
  AUTHORS   Roemer,T., Jiang,B., Boone,C. and Bussey,H.
  TITLE     Gene disruption methodologies for drug target discovery
  JOURNAL   Patent: WO 0160975-A 7 23-AUG-2001;
            Elitra Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .1521
                     /organism="Candida albicans"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:5476"
ORIGIN

Alignment Scores:
Pred. No.:              0.351          Length:          1521
Score:                  76.00          Matches:         11
Percent Similarity:     100.00%        Conservative:    0
Best Local Similarity:  100.00%        Mismatches:      0
Query Match:            90.48%         Indels:          0
DB:                     6              Gaps:            0

US-09-858-332G-15 (1-12) x AX240769 (1-1521)

Qy          2 AsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
                ||||||||||||||||||||||||||||||||
Db        173 AACCACAACCACAACCACAACCACAACCACAAC 205

RESULT 208
AX488745
LOCUS       AX488745               1521 bp    DNA     linear   PAT 16-AUG-2002
DEFINITION  Sequence 6045 from Patent WO02053728.
ACCESSION   AX488745
VERSION     AX488745.1  GI:22322757
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1
  AUTHORS   Roemer,T., Jiang,B., Boone,C., Bussey,H. and Ohlsen,K.L.
  TITLE     Gene disruption methodologies for drug target discovery
  JOURNAL   Patent: WO 02053728-A 6045 11-JUL-2002;
            Elitra Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .1521
                     /organism="Candida albicans"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:5476"
ORIGIN

Alignment Scores:
Pred. No.:              0.351          Length:          1521
Score:                  76.00          Matches:         11
Percent Similarity:     100.00%        Conservative:    0
Best Local Similarity:  100.00%        Mismatches:      0
Query Match:            90.48%         Indels:          0
DB:                     6              Gaps:            0

US-09-858-332G-15 (1-12) x AX488745 (1-1521)

Qy          2 AsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
                ||||||||||||||||||||||||||||||||
Db        173 AACCACAACCACAACCACAACCACAACCACAAC 205

RESULT 209
AR545053
LOCUS       AR545053               1806 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 184 from patent US 6747137.
ACCESSION   AR545053
VERSION     AR545053.1  GI:53938228
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1806)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 184 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .1806
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              0.414          Length:          1806
Score:                  76.00          Matches:         11
Percent Similarity:     100.00%        Conservative:    0
Best Local Similarity:  100.00%        Mismatches:      0
Query Match:            90.48%         Indels:          0
DB:                     6              Gaps:            0

US-09-858-332G-15 (1-12) x AR545053 (1-1806)

Qy          2 AsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
                ||||||||||||||||||||||||||||||||
Db       1147 AACCATAACCACAACCATAACCACAACCACAAC 1179

RESULT 210
AF332954
LOCUS       AF332954               3897 bp    DNA     linear   PLN 03-JUN-2005
DEFINITION  Zea mays transposon Ds2(D4), complete sequence.
ACCESSION   AF332954
VERSION     AF332954.1  GI:13345975
KEYWORDS    .
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 3897)
  AUTHORS   Yan,X., Martinez-Ferez,I.M., Kavchok,S. and Dooner,H.K.
  TITLE     Origination of Ds elements from Ac elements in maize: evidence for
            rare repair synthesis at the site of Ac excision
  JOURNAL   Genetics 152 (4), 1733-1740 (1999)
   PUBMED   10430597
REFERENCE   2  (bases 1 to 3897)
  AUTHORS   Yan,X., Martinez-Ferez,I.M., Kavchok,S. and Dooner,H.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JAN-2001) Waksman Institute, Rutgers University,
            Piscataway, NJ 08855, USA
FEATURES             Location/Qualifiers
     source          1. .3897
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4577"
                     /chromosome="9"
                     /map="9S"
                     /note="internal deletion derivative 4 of Ac element in
                     bz-m2(Ac)"
     repeat_region   1. .3897
```

```
                     /note="internal deletion derivative 1 of Ac element in
                     bz-m2(Ac)"
                     /transposon="Ds2(D4)"
ORIGIN

Alignment Scores:
Pred. No.:                0.869          Length:         3897
Score:                    76.00          Matches:        11
Percent Similarity:       91.67%         Conservative:   0
Best Local Similarity:    91.67%         Mismatches:     1
Query Match:              90.48%         Indels:         0
DB:                       15             Gaps:           0

US-09-858-332G-15 (1-12) x AF332954 (1-3897)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||   |||||||||
Db       1420 CACAACCACAACCACAACCACAACCAGAACCACAAC 1455

RESULT 211
AF332955
LOCUS       AF332955                4167 bp    DNA     linear   PLN 03-JUN-2005
DEFINITION  Zea mays transposon Ds2(D5), complete sequence.
ACCESSION   AF332955
VERSION     AF332955.1  GI:13345976
KEYWORDS    .
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 4167)
  AUTHORS   Yan,X., Martinez-Ferez,I.M., Kavchok,S. and Dooner,H.K.
  TITLE     Origination of Ds elements from Ac elements in maize: evidence for
            rare repair synthesis at the site of Ac excision
  JOURNAL   Genetics 152 (4), 1733-1740 (1999)
   PUBMED   10430597
REFERENCE   2  (bases 1 to 4167)
  AUTHORS   Yan,X., Martinez-Ferez,I.M., Kavchok,S. and Dooner,H.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JAN-2001) Waksman Institute, Rutgers University,
            Piscataway, NJ 08855, USA
FEATURES             Location/Qualifiers
     source          1. .4167
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4577"
                     /chromosome="9"
                     /map="9S"
                     /note="internal deletion derivative 5 of Ac element in
                     bz-m2(Ac)"
     repeat_region   1. .4167
                     /note="internal deletion derivative 1 of Ac element in
                     bz-m2(Ac)"
                     /transposon="Ds2(D5)"
ORIGIN

Alignment Scores:
Pred. No.:                0.927          Length:         4167
Score:                    76.00          Matches:        11
Percent Similarity:       91.67%         Conservative:   0
Best Local Similarity:    91.67%         Mismatches:     1
Query Match:              90.48%         Indels:         0
DB:                       15             Gaps:           0

US-09-858-332G-15 (1-12) x AF332955 (1-4167)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||   |||||||||
Db       1420 CACAACCACAACCACAACCACAACCAGAACCACAAC 1455

RESULT 212
DD013389/c
LOCUS       DD013389                4565 bp    DNA     linear   PAT 02-FEB-2005
DEFINITION  Methods and Compositions to Reduce or Eliminate Transmission of a
            Transgene.
ACCESSION   DD013389
VERSION     DD013389.1  GI:58530792
KEYWORDS    JP 2004512007-A/3.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 4565)
  AUTHORS   Dellaporta,S.L. and Moreno,M.A.
  TITLE     Methods and Compositions to Reduce or Eliminate Transmission of a
            Transgene
  JOURNAL   Patent: JP 2004512007-A 3 22-APR-2004;
            Stephen L Dellaporta,Maria A Moreno,Yale University
COMMENT     OS   Zea mays
            PN   JP 2004512007-A/3
            PD   22-APR-2004
            PF   28-FEB-2001 JP 2001563615
            PR   28-FEB-2000 US    60/185524
            PI   stephen l dellaporta,maria a moreno
            CC   Transposable element Ac
            FH   Key             Location/Qualifiers.
FEATURES             Location/Qualifiers
     source          1. .4565
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4577"
ORIGIN

Alignment Scores:
Pred. No.:                1.01           Length:         4565
Score:                    76.00          Matches:        11
Percent Similarity:       91.67%         Conservative:   0
Best Local Similarity:    91.67%         Mismatches:     1
Query Match:              90.48%         Indels:         0
DB:                       6              Gaps:           0

US-09-858-332G-15 (1-12) x DD013389 (1-4565)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||   |||||||||
Db       3146 CACAACCACAACCACAACCACAACCAGAACCACAAC 3111

RESULT 213
AR542626/c
LOCUS       AR542626                4565 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 3 from patent US 6743968.
ACCESSION   AR542626
VERSION     AR542626.1  GI:53935114
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 4565)
  AUTHORS   Dellaporta,S.L. and Moreno,M.A.
  TITLE     Methods and compositions to reduce or eliminate transmission of a
            transgene
  JOURNAL   Patent: US 6743968-A 3 01-JUN-2004;
            Yale University; New Haven, CT
FEATURES             Location/Qualifiers
     source          1. .4565
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:                1.01           Length:         4565
Score:                    76.00          Matches:        11
Percent Similarity:       91.67%         Conservative:   0
Best Local Similarity:    91.67%         Mismatches:     1
```

```
Query Match:            90.48%        Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-15 (1-12) x AR542626 (1-4565)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||   |||||||||
Db       3146 CACAACCACAACCACAACCACAACCAGAACCACAAC 3111

RESULT 214
AX235897/c
LOCUS       AX235897               4565 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 3 from Patent WO0164926.
ACCESSION   AX235897
VERSION     AX235897.1  GI:15795786
KEYWORDS    .
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1
  AUTHORS   Dellaporta,S.L. and Moreno,M.A.
  TITLE     Methods and compositions to reduce or eliminate transmission of a
            transgene
  JOURNAL   Patent: WO 0164926-A 3 07-SEP-2001;
            YALE UNIVERSITY (US)
FEATURES             Location/Qualifiers
     source          1. .4565
                     /organism="Zea mays"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:4577"
                     /note="Transposable element Ac"
ORIGIN

Alignment Scores:
Pred. No.:              1.01          Length:        4565
Score:                  76.00         Matches:       11
Percent Similarity:     91.67%        Conservative:  0
Best Local Similarity:  91.67%        Mismatches:    1
Query Match:            90.48%        Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-15 (1-12) x AX235897 (1-4565)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||   |||||||||
Db       3146 CACAACCACAACCACAACCACAACCAGAACCACAAC 3111

RESULT 215
ZMAC1
LOCUS       ZMAC1                  4565 bp    mRNA    linear   PLN 10-FEB-1999
DEFINITION  Maize transposable element Activator (Ac) major transcript.
ACCESSION   X05424 X05425
VERSION     X05424.1  GI:22112
KEYWORDS    Ac element; transposable element.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1
  AUTHORS   Kunze,R., Stochaj,U., Laufs,J. and Starlinger,P.
  TITLE     Transcription of transposable element Activator (Ac) of Zea mays L
  JOURNAL   EMBO J. 6, 1555-1563 (1987)
REFERENCE   2  (bases 1 to 4565)
  AUTHORS   Kunze,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-FEB-1988) R. Kunze, Institut fuer Genetik,
            Universitaet zu Koeln, D-5000 Koeln 41, West Germany
FEATURES             Location/Qualifiers
     source          1. .4565
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="taxon:4577"
     misc_feature    303
                     /note="minor transcription start site"
     misc_feature    334
                     /note="major transcription start site"
     misc_feature    343
                     /note="minor transcription start site"
     misc_feature    359
                     /note="minor transcription start site"
     misc_feature    363
                     /note="minor transcription start site"
     misc_feature    569. .2235
                     /note="pcAcP"
     CDS             join(988. .1111,1219. .2566,2638. .3434,3524. .3615,
                     4003. .4065)
                     /codon_start=1
                     /product="ORFa"
                     /protein_id="CAA29005.1"
                     /db_xref="GI:22113"
                     /db_xref="GOA:P08770"
                     /db_xref="UniProt/Swiss-Prot:P08770"
                     /translation="MTPPVGNNPPSGSAIRLAKLMSTTRAPSTRKTNSVFSAYAQGLK
                     RKAEASSSRIQNVRARARGHGCGRTSPSSSTAEAERHFIQSVSSSNANGTATDPSQDD
                     MAIVHEPQPQPQPQPEPQPQPQPEPEEEAPQKRAKKCTSDVWQHFTKKEIEVEVDGKK
                     YVQVWGHCNFPNCKAKYRAEGHHGTSGFRNHLRTSHSLVKGQLCLKSEKDHGKDINLI
                     EPYKYDEVVSLKKLHLAIIMHEYPFNIVEHEYFVEFVKSLRPHFPIKSRVTARKYIMD
                     LYLEEKEKLYGKLKDVQSRFSTTMDMWTSCQNKSYMCVTIHWIDDDWCLQKRIVGFFH
                     VEGRHTGQRLSQTFTAIMVKWNIEKKLFALSLDNASANEVAVHDIIEDLQDTDSNLVC
                     DGAFFHVRCACHILNLVAKDGLAVIAGTIEKIKAIVLAVKSSPLQWEELMKCASECDL
                     DKSKGISYDVSTRWNSTYLMLRDALYYKPALIRLKTSDPRRYDAICPKAEEWKMALTL
                     FKCLKKFFDLTELLSGTQYSTANLFYKGFCEIKDLIDQWCVHEKFVIRRMAVAMSEKF
                     EKYWKVSNIALAVACFLDPRYKKILIEFYMKKFHGDSYKVHVDDFVRVIRKLYQFYSS
                     CSPSAPKTKTTTNDSMDDTLMENEDDEFQNYLHELKDYDQVESNELDKYMSEPLLKHS
                     GQFDILSWWRGRVAEYPILTQIARDVLAIQVSTVASESAFSAGGRVVDPYRNRLGSEI
                     VEALICTKDWVAASRKGATYFPTMIGDLEVLDSVIAAATNHENHMDEDEDAIEFSKNN
                     EDVASGSSP"
     intron          1112. .1218
                     /number=1
     CDS             1363. .1671
                     /codon_start=1
                     /product="ORFb"
                     /protein_id="CAA29006.1"
                     /db_xref="GI:22114"
                     /db_xref="GOA:P08771"
                     /db_xref="UniProt/Swiss-Prot:P08771"
                     /translation="MQMVQLQIRVKMIWLLFMNHNHNHNHNQNHNHSHNLNPKKKHHR
                     RGQRSAHRMYGSISPRRKLKWRSMERNTFRYGDIATFLIARLSIGLRVIMEQADFEIT
                     "
     misc_feature    1589. .4246
                     /note="pcAcXY"
     intron          2567. .2637
                     /number=2
     misc_feature    2754. .4302
                     /note="pcAcQ"
     intron          3435. .3523
                     /number=3
     intron          3616. .4002
                     /number=4
     polyA_site      4302
ORIGIN

Alignment Scores:
Pred. No.:              1.01          Length:        4565
Score:                  76.00         Matches:       11
Percent Similarity:     91.67%        Conservative:  0
Best Local Similarity:  91.67%        Mismatches:    1
Query Match:            90.48%        Indels:        0
DB:                     15            Gaps:          0

US-09-858-332G-15 (1-12) x ZMAC1 (1-4565)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
```

```
                    ||||||||||||||||||||||||  |||||||||
Db       1420 CACAACCACAACCACAACCACAACCAGAACCACAAC 1455

RESULT 216
ZMTAC/c
LOCUS       ZMTAC                    4565 bp    DNA     linear   PLN 07-JUL-2002
DEFINITION  Zea mays transposable element Ac.
ACCESSION   X01380
VERSION     X01380.1  GI:22489
KEYWORDS    transposon; unidentified reading frame.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 4565)
  AUTHORS   Mueller-Neumann,M., Yoder,J.I. and Starlinger,P.
  TITLE     The DNA sequence of the transposable element Ac of Zea mays L
  JOURNAL   Mol. Gen. Genet. 198, 19-24 (1984)
REFERENCE   2
  AUTHORS   Kunze,R. and Starlinger,P.
  TITLE     The putative transposase of transposable element Ac from Zea mays
            L. interacts with subterminal sequences of Ac
  JOURNAL   EMBO J. 8 (11), 3177-3185 (1989)
   PUBMED   2555157
REFERENCE   3  (bases 1 to 4565)
  AUTHORS   Kunze,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAR-1986)
COMMENT     Data kindly reviewed (20-AUG-1985) by M. Mueller-Neumann.
FEATURES             Location/Qualifiers
     source          1. .4565
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4577"
     repeat_region   1. .4565
                     /transposon="Ac element"
     repeat_unit     1. .11
                     /note="imperfect inverted repeat A"
     polyA_signal    complement(992. .996)
                     /note="ORF1"
     polyA_signal    complement(1021. .1026)
                     /note="ORF1"
     CDS             complement(1064. .1729)
                     /note="unnamed protein product; ORF1"
                     /codon_start=1
                     /protein_id="CAA25635.1"
                     /db_xref="GI:22490"
                     /db_xref="UniProt/TrEMBL:Q41857"
                     /translation="MAVAMSEKFEKYWKVSNIALAVACFLDPRYKKILIEFYMKKFHG
                     DSYKVHVDDFVRVIRKLYQFYSSCSPSAPKTKTTTNDSMDDTLMENEDDEFQNYLHEL
                     KDYDQVESNELDKYMSEPLLKHSGQFDILSWWRGRVAEYPILTQIARDVLAIQVSTVA
                     SESAFSAGGRVVDPYRNRLGSEIVEALICTKDWVAASRKGECIYVIMKFQFIVIQQLF
                     YLY"
     polyA_signal    complement(1750. .1753)
                     /note="ORF2"
     polyA_signal    complement(1761. .1765)
                     /note="ORF2"
     polyA_signal    complement(1779. .1784)
                     /note="ORF2"
     TATA_signal     complement(1800. .1806)
     CDS             complement(1882. .3165)
                     /note="unnamed protein product; ORF2"
                     /codon_start=1
                     /protein_id="CAA25636.1"
                     /db_xref="GI:22491"
                     /db_xref="GOA:Q9SAP9"
                     /db_xref="UniProt/TrEMBL:Q9SAP9"
                     /translation="MAIVHEPQPQPQPQPEPQPQPQPEPEEEAPQKRAKKCTSDVWQH
                     FTKKEIEVEVDGKKYVQVWGHCNFPNCKAKYRAEGHHGTSGFRNHLRTSHSLVKGQLC
                     LKSEKDHGKDINLIEPYKYDEVVSLKKLHLAIIMHEYPFNIVEHEYFVEFVKSLRPHF
                     PIKSRVTARKYIMDLYLEEKEKLYGKLKDVQSRFSTTMDMWTSCQNKSYMCVTIHWID
                     DDWCLQKRIVGFFHVEGRHTGQRLSQTFTAIMVKWNIEKKLFALSLDNASANEVAVHD
                     IIEDLQDTDSNLVCDGAFFHVRCACHILNLVAKDGLAVIAGTIEKIKAIVLAVKSSPL
                     QWEELMKCASECDLDKSKGISYDVSTRWNSTYLMLRDALYYKPALIRLKTSDPRRYVC
                     LNCCTCHHYKFSINQMSIIVGTMQFVLKPRSGRWH"
     TATA_signal     complement(3221. .3228)
                     /note="ORF3
                     putative"
     TATA_signal     3557. .3565
                     /note="ORF3
                     putative"
     CDS             3739. .4203
                     /note="unnamed protein product; ORF3"
                     /codon_start=1
                     /protein_id="CAA25637.1"
                     /db_xref="GI:1666226"
                     /db_xref="UniProt/TrEMBL:Q41858"
                     /translation="MTQQLGAMVLTCCTAKCCEPVSSRGDRETADGRMGERRAVEVWR
                     TADRRWRMADGRTADGRAVEWRSGRMGGPRRGWASREQWSGGRTVDGGVCVRRAASRP
                     SPRVVPGAAQRPAGWETGRVGESRRERDASPASACGWRRPDSGVGAWRRVNV"
     polyA_signal    4496. .4501
                     /note="ORF3
                     putative"
     polyA_signal    4522. .4526
                     /note="ORF3
                     putative"
     repeat_unit     4555. .4565
                     /note="imperfect direct repeat A'"
ORIGIN

Alignment Scores:
Pred. No.:                1.01          Length:          4565
Score:                    76.00         Matches:         11
Percent Similarity:       91.67%        Conservative:    0
Best Local Similarity:    91.67%        Mismatches:      1
Query Match:              90.48%        Indels:          0
DB:                       15            Gaps:            0

US-09-858-332G-15 (1-12) x ZMTAC (1-4565)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||  |||||||||
Db       3146 CACAACCACAACCACAACCACAACCAGAACCACAAC 3111

RESULT 217
MZEAC9WX/c
LOCUS       MZEAC9WX                 4810 bp    DNA     linear   PLN 28-FEB-1994
DEFINITION  Maize transposable element Activator (Ac9) from the waxy locus.
ACCESSION   K01964
VERSION     K01964.1  GI:168402
KEYWORDS    activator; repeat region; transposase; transposon.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 4810)
  AUTHORS   Pohlman,R.F., Fedoroff,N.V. and Messing,J.
  TITLE     The nucleotide sequence of the maize controlling element Activator
  JOURNAL   Cell 37 (2), 635-643 (1984)
   PUBMED   6327080
REFERENCE   2  (bases 1 to 4810)
  AUTHORS   Pohlman,R.F., Fedoroff,N.V. and Messing,J.
  TITLE     Correction: nucleotide sequence of Ac
  JOURNAL   Cell 39 (2 Pt 1), 417 (1984)
   PUBMED   6094018
COMMENT     Original source text: Maize DNA, clone Ac wx-m9.
            [2]  revises [1].
            The ends of the Ac9 element comprise an 11 bp imperfect terminal
            repetition and the insertion site within the wx locus is flanked by
            an 8 bp directly repeated sequence.
FEATURES             Location/Qualifiers
     source          1. .4810
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
```

```
                     /db_xref="taxon:4577"
     misc_feature    114. .121
                     /note="insertion target sequence copy A"
     misc_feature    122. .4682
                     /standard_name="transposable element Ac9"
     repeat_region   122. .132
                     /note="inverted repeat copy A (imperfect)"
     repeat_region   4672. .4682
                     /note="inverted repeat copy B (imperfect)"
     misc_feature    4683. .4690
                     /note="insertion target sequence copy B"
ORIGIN      5 bp upstream of PstI site.

Alignment Scores:
Pred. No.:              1.06          Length:         4810
Score:                  76.00         Matches:        11
Percent Similarity:     91.67%        Conservative:   0
Best Local Similarity:  91.67%        Mismatches:     1
Query Match:            90.48%        Indels:         0
DB:                     15            Gaps:           0

US-09-858-332G-15 (1-12) x MZEAC9WX (1-4810)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||   |||||||||
Db       3267 CACAACCACAACCACAACCACAACCAGAACCACAAC 3232

RESULT 218
AF355378
LOCUS       AF355378               5772 bp    DNA     linear   PLN 01-MAY-2001
DEFINITION  Zea mays bz-m2(Ac) mutable allele, genomic sequence.
ACCESSION   AF355378
VERSION     AF355378.1  GI:13898794
KEYWORDS    .
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 5772)
  AUTHORS   Fu,H. and Dooner,H.K.
  TITLE     The Ac element in the bz-m2(Ac) allele from maize
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 5772)
  AUTHORS   Fu,H. and Dooner,H.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2001) Waksman Institute, Rutgers University,
            Piscataway, NJ 08854, USA
FEATURES             Location/Qualifiers
     source          1. .5772
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4577"
                     /chromosome="9"
                     /map="9S"
     source          525. .5089
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4577"
                     /chromosome="9"
                     /map="9S"
                     /transposon="Activator (Ac)"
                     /note="Ac inserted into the Bz-McC allele"
     gene            <1. .>5772
                     /gene="bz"
                     /allele="bz-m2(Ac)"
     misc_feature    517. .524
                     /gene="bz"
                     /allele="bz-m2(Ac)"
                     /note="duplicated integration site sequence"
     misc_feature    5090. .5097
                     /gene="bz"
                     /allele="bz-m2(Ac)"
                     /note="duplicated integration site sequence"
ORIGIN

Alignment Scores:
Pred. No.:              1.27          Length:         5772
Score:                  76.00         Matches:        11
Percent Similarity:     91.67%        Conservative:   0
Best Local Similarity:  91.67%        Mismatches:     1
Query Match:            90.48%        Indels:         0
DB:                     15            Gaps:           0

US-09-858-332G-15 (1-12) x AF355378 (1-5772)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||   |||||||||
Db       1944 CACAACCACAACCACAACCACAACCAGAACCACAAC 1979

RESULT 219
AF433042
LOCUS       AF433042               6875 bp    DNA     circular SYN 11-NOV-2001
DEFINITION  Cloning vector pWS31, complete sequence.
ACCESSION   AF433042
VERSION     AF433042.1  GI:16903089
KEYWORDS    .
SOURCE      Cloning vector pWS31
  ORGANISM  Cloning vector pWS31
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 6875)
  AUTHORS   Sundaresan,V., Springer,P., Volpe,T., Haward,S., Jones,J.D.,
            Dean,C., Ma,H. and Martienssen,R.
  TITLE     Patterns of gene action in plant development revealed by enhancer
            trap and gene trap transposable elements
  JOURNAL   Genes Dev. 9 (14), 1797-1810 (1995)
   PUBMED   7622040
REFERENCE   2  (bases 1 to 6875)
  AUTHORS   de la Bastide,M., Preston,R., O'Shaughnessy,A., May,B.,
            Sundaresan,V., Martienssen,R. and McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-OCT-2001) Cold Spring Harbor Laboratory, 1 Bungtown
            Rd, Cold Spring Harbor, NY 11724, USA
FEATURES             Location/Qualifiers
     source          1. .6875
                     /organism="Cloning vector pWS31"
                     /mol_type="other DNA"
                     /db_xref="taxon:176659"
     source          64. .1852
                     /organism="Cloning vector pWS31"
                     /mol_type="other DNA"
                     /db_xref="taxon:176659"
                     /transposon="Ds transposable element"
                     /note="5' fragment"
     source          6458. .6649
                     /organism="Cloning vector pWS31"
                     /mol_type="other DNA"
                     /db_xref="taxon:176659"
                     /transposon="Ds transposable element"
                     /note="3' fragment; transferred to T-DNA vector pAJ8 for
                     transformation of Arabidopsis thaliana"
     source          6670. .6861
                     /organism="Cloning vector pWS31"
                     /mol_type="other DNA"
                     /db_xref="taxon:176659"
                     /transposon="Ds transposable element"
                     /note="3' fragment"
     gene            2376. .3170
                     /gene="npt2"
     CDS             2376. .3170
                     /gene="npt2"
                     /codon_start=1
                     /transl_table=11
                     /product="neomycin phosphotransferase"
                     /protein_id="AAL30400.1"
                     /db_xref="GI:16903090"
```

```
                     /translation="MIEQDGLHAGSPAAWVERLFGYDWAQQTIGCSDAAVFRLSAQGR
                     PVLFVKTDLSGALNELQDEAARLSWLATTGVPCAAVLDVVTEAGRDWLLLGEVPGQDL
                     LSSHLAPAEKVSIRLNAMRRLHTLDPATCPFDHQAKHRIERARTRMEAGLVDQDDLDE
                     EHQGLAPAELFARLKARMPDGEDLVVTHGDACLPNIMVENGRFSGFIDCGRLGVADRY
                     QDIALATRDIAEELGGEWADRFLVLYGIAAPDSQRIAFYRLLDEFF"
     gene            complement(4550. .6361)
                     /gene="uidA"
     CDS             complement(4550. .6361)
                     /gene="uidA"
                     /codon_start=1
                     /transl_table=11
                     /product="beta-glucuronidase"
                     /protein_id="AAL30401.1"
                     /db_xref="GI:16903091"
                     /translation="MVRPVETPTREIKKLDGLWAFSLDRENCGIDQRWWESALQESRA
                     IAVPGSFNDQFADADIRNYAGNVWYQREVFIPKGWAGQRIVLRFDAVTHYGKVWVNNQ
                     EVMEHQGGYTPFEADVTPYVIAGKSVRITVCVNNELNWQTIPPGMVITDENGKKKQSY
                     FHDFFNYAGIHRSVMLYTTPNTWVDDITVVTHVAQDCNHASVDWQVVANGDVSVELRD
                     ADQQVVATGQGTSGTLQVVNPHLWQPGEGYLYELCVTAKSQTECDIYPLRVGIRSVAV
                     KGEQFLINHKPFYFTGFGRHEDADLRGKGFDNVLMVHDHALMDWIGANSYRTSHYPYA
                     EEMLDWADEHGIVVIDETAAVGFNLSLGIGFEAGNKPKELYSEEAVNGETQQAHLQAI
                     KELIARDKNHPSVVMWSIANEPDTRPQGAREYFAPLAEATRKLDPTRPITCVNVMFCD
                     AHTDTISDLFDVLCLNRYYGWYVQSGDLETAEKVLEKELLAWQEKLHQPIIITEYGVD
                     TLAGLHSMYTDMWSEEYQCAWLDMYHRVFDRVSAVVGEQVWNFADFATSQGILRVGGN
                     KKGIFTRDRKPKSAAFLLQKRWTGMNFGEKPQQGGKQ"
     misc_feature    complement(6381. .6414)
                     /note="CaMV 35S promoter -46 region"
ORIGIN

Alignment Scores:
Pred. No.:                1.5          Length:        6875
Score:                    76.00        Matches:       11
Percent Similarity:       91.67%       Conservative:  0
Best Local Similarity:    91.67%       Mismatches:    1
Query Match:              90.48%       Indels:        0
DB:                       11           Gaps:          0

US-09-858-332G-15 (1-12) x AF433042 (1-6875)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||   |||||||||
Db       1483 CACAACCACAACCACAACCACAACCAGAACCACAAC 1518

RESULT 220
AF433043
LOCUS       AF433043                6928 bp    DNA     circular SYN 11-NOV-2001
DEFINITION  Cloning vector pWS32, complete sequence.
ACCESSION   AF433043
VERSION     AF433043.1  GI:16903092
KEYWORDS    .
SOURCE      Cloning vector pWS32
  ORGANISM  Cloning vector pWS32
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 6928)
  AUTHORS   Sundaresan,V., Springer,P., Volpe,T., Haward,S., Jones,J.D.,
            Dean,C., Ma,H. and Martienssen,R.
  TITLE     Patterns of gene action in plant development revealed by enhancer
            trap and gene trap transposable elements
  JOURNAL   Genes Dev. 9 (14), 1797-1810 (1995)
   PUBMED   7622040
REFERENCE   2  (bases 1 to 6928)
  AUTHORS   de la Bastide,M., Preston,R., O'Shaughnessy,A., May,B.,
            Sundaresan,V., Martienssen,R. and McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-OCT-2001) Cold Spring Harbor Laboratory, 1 Bungtown
            Rd, Cold Spring Harbor, NY 11724, USA
FEATURES             Location/Qualifiers
     source          1. .6928
                     /organism="Cloning vector pWS32"
                     /mol_type="other DNA"
                     /db_xref="taxon:176660"
     source          64. .1852
                     /organism="Cloning vector pWS32"
                     /mol_type="other DNA"
                     /db_xref="taxon:176660"
                     /transposon="Ds transposable element"
                     /note="5' fragment"
     source          6511. .6702
                     /organism="Cloning vector pWS32"
                     /mol_type="other DNA"
                     /db_xref="taxon:176660"
                     /transposon="Ds transposable element"
                     /note="3' fragment; transferred to T-DNA vector pAJ8 for
                     transformation of Arabidopsis thaliana"
     source          6720. .6911
                     /organism="Cloning vector pWS32"
                     /mol_type="other DNA"
                     /db_xref="taxon:176660"
                     /transposon="Ds transposable element"
                     /note="3' fragment"
     gene            2376. .3170
                     /gene="npt2"
     CDS             2376. .3170
                     /gene="npt2"
                     /codon_start=1
                     /transl_table=11
                     /product="neomycin phosphotransferase"
                     /protein_id="AAL30402.1"
                     /db_xref="GI:16903093"
                     /translation="MIEQDGLHAGSPAAWVERLFGYDWAQQTIGCSDAAVFRLSAQGR
                     PVLFVKTDLSGALNELQDEAARLSWLATTGVPCAAVLDVVTEAGRDWLLLGEVPGQDL
                     LSSHLAPAEKVSIMADAMRRLHTLDPATCPFDHQAKHRIERARTRMEAGLVDQDDLDE
                     EHQGLAPAELFARLKARMPDGEDLVVTHGDACLPNIMVENGRFSGFIDCGRLGVADRY
                     QDIALATRDIAEELGGEWADRFLVLYGIAAPDSQRIAFYRLLDEFF"
     gene            complement(4550. .6361)
                     /gene="uidA"
     CDS             complement(4550. .6361)
                     /gene="uidA"
                     /codon_start=1
                     /transl_table=11
                     /product="beta-glucuronidase"
                     /protein_id="AAL30403.1"
                     /db_xref="GI:16903094"
                     /translation="MVRPVETPTREIKKLDGLWAFSLDRENCGIDQRWWESALQESRA
                     IAVPGSFNDQFADADIRNYAGNVWYQREVFIPKGWAGQRIVLRFDAVTHYGKVWVNNQ
                     EVMEHQGGYTPFEADVTPYVIAGKSVRITVCVNNELNWQTIPPGMVITDENGKKKQSY
                     FHDFFNYAGIHRSVMLYTTPNTWVDDITVVTHVAQDCNHASVDWQVVANGDVSVELRD
                     ADQQVVATGQGTSGTLQVVNPHLWQPGEGYLYELCVTAKSQTECDIYPLRVGIRSVAV
                     KGEQFLINHKPFYFTGFGRHEDADLRGKGFDNVLMVHDHALMDWIGANSYRTSHYPYA
                     EEMLDWADEHGIVVIDETAAVGFNLSLGIGFEAGNKPKELYSEEAVNGETQQAHLQAI
                     KELIARDKNHPSVVMWSIANEPDTRPQGAREYFAPLAEATRKLDPTRPITCVNVMFCD
                     AHTDTISDLFDVLCLNRYYGWYVQSGDLETAEKVLEKELLAWQEKLHQPIIITEYGVD
                     TLAGLHSMYTDMWSEEYQCAWLDMYHRVFDRVSAVVGEQVWNFADFATSQGILRVGGN
                     KKGIFTRDRKPKSAAFLLQKRWTGMNFGEKPQQGGKQ"
     intron          complement(6364. .6615)
                     /note="synthetic GPA1 intron containing splice donors and
                     acceptors in 3 reading frames"
ORIGIN

Alignment Scores:
Pred. No.:                1.51         Length:        6928
Score:                    76.00        Matches:       11
Percent Similarity:       91.67%       Conservative:  0
Best Local Similarity:    91.67%       Mismatches:    1
Query Match:              90.48%       Indels:        0
DB:                       11           Gaps:          0

US-09-858-332G-15 (1-12) x AF433043 (1-6928)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||   |||||||||
Db       1483 CACAACCACAACCACAACCACAACCAGAACCACAAC 1518

RESULT 221
AB055064
LOCUS       AB055064                7635 bp    DNA     linear   SYN 21-MAY-2003
```

```
DEFINITION  Synthetic transposable element dAc-I-RS DNA, complete sequence.
ACCESSION   AB055064
VERSION     AB055064.1  GI:13359266
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Machida,C., Onouchi,H., Koizumi,J., Hamada,S., Semiarti,E.,
            Torikai,S. and Machida,Y.
  TITLE     Characterization of the transposition pattern of the Ac element in
            Arabidopsis thaliana using endonuclease I-SceI
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 94 (16), 8675-8680 (1997)
   PUBMED   11038561
REFERENCE   2
  AUTHORS   Semiarti,E., Onouchi,H., Torikai,S., Ishikawa,T., Machida,Y. and
            Machida,C.
  TITLE     The transposition pattern of the Ac element in tobacco cultured
            cells
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 7635)
  AUTHORS   Machida,C. and Machida,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JAN-2001) Yasunori Machida, Nagoya University,
            Division of Biological Science, Graduate School of Science;
            Chikusa-ku, Furo-cho, Nagoya, Aichi 464-8602, Japan
            (E-mail:yas@bio.nagoya-u.ac.jp, Tel:81-52-789-5040,
            Fax:81-52-789-2966)
FEATURES             Location/Qualifiers
     source          1. .7635
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="non-autonomous derivative of Ac
                     transposable element dAc-I-RS"
     misc_feature    1. .1789
                     /note="sequence of the Ac element in maize, containing the
                     terminal inverted repeat"
     misc_feature    1807. .1864
                     /note="RS recombination site recognized by R protein from
                     Zygosaccharomyces rouxii"
     promoter        2016. .2380
                     /note="the 35S promoter from cauliflower mosaic virus"
     gene            2390. .3415
                     /gene="HPT"
     CDS             2390. .3415
                     /gene="HPT"
                     /codon_start=1
                     /transl_table=11
                     /product="hygromycin phosphotransferase"
                     /protein_id="BAB33367.1"
                     /db_xref="GI:13359267"
                     /translation="MKKPELTATSVEKFLIEKFDSVSDLMQLSEGEESRAFSFDVGGR
                     GYVLRVNSCADGFYKDRYVYRHFASAALPIPEVLDIGEFSESLTYCISRRAQGVTLQD
                     LPETELPAVLQPVAEAMDAIAAADLSQTSGFGPFGPQGIGQYTTWRDFICAIADPHVY
                     HWQTVMDDTVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTDNGRITAVIDWS
                     EAMFGDSQYEVANIFFWRPWLACMEQQTRYFERRHPELAGSPRLRAYMLRIGLDQLYQ
                     SLVDGNFDDAAWAQGRCDAIVRSGAGTVGRTQIARRSAAVWTDGCVEVLADSGNRRPS
                     TRPRAKE"
     terminator      3436. .4560
                     /note="the terminator of the gene for nopaline synthase"
     misc_feature    4575. .4632
                     /note="RS recombination site recognized by R protein from
                     Zygosaccharomyces rouxii"
     misc_feature    4683. .4700
                     /note="cleavage site for endonuclease I-SceI"
     misc_feature    5149. .5172
                     /note="cos sequence"
     misc_feature    6496. .7635
                     /note="sequence of the Ac element in maize, containing the
                     terminal inverted repeat"
ORIGIN
```

```
Alignment Scores:
Pred. No.:              1.66          Length:          7635
Score:                  76.00         Matches:         11
Percent Similarity:     91.67%        Conservative:    0
Best Local Similarity:  91.67%        Mismatches:      1
Query Match:            90.48%        Indels:          0
DB:                     11            Gaps:            0

US-09-858-332G-15 (1-12) x AB055064 (1-7635)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||||||||||||||||||||||   |||||||||
Db       1420 CACAACCACAACCACAACCACAACCAGAACCACAAC 1455

RESULT 222
CDR131945
LOCUS       CDR131945               8724 bp    DNA     linear   MAM 07-JAN-2000
DEFINITION  Camelus dromedarius IGG2a gene, alternative transcripts.
ACCESSION   AJ131945
VERSION     AJ131945.1  GI:6687180
KEYWORDS    alternative splicing; constant region; IGG2a gene; immunoglobulin
            heavy chain; membrane-bound form; secreted form.
SOURCE      Camelus dromedarius (Arabian camel)
  ORGANISM  Camelus dromedarius
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Tylopoda;
            Camelidae; Camelus.
REFERENCE   1
  AUTHORS   Nguyen,V.K., Hamers,R., Wyns,L. and Muyldermans,S.
  TITLE     Loss of splice consensus signal is responsible for the removal of
            the entire C(H)1 domain of the functional camel IGG2A heavy-chain
            antibodies
  JOURNAL   Mol. Immunol. 36 (8), 515-524 (1999)
   PUBMED   10475606
REFERENCE   2  (bases 1 to 8724)
  AUTHORS   Nguyen,K.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-1999) Nguyen K.V., ULTR, Insitute of
            biotechnology, Vrije Universiteit Brussel, Paardenstraat 65, Sint
            Genesius Rode, B-1640, BELGIUM
FEATURES             Location/Qualifiers
     source          1. .8724
                     /organism="Camelus dromedarius"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9838"
                     /clone="ch51666E2"
                     /tissue_type="liver"
                     /germline
     gene            1. .3863
                     /gene="IGG2a"
     5'UTR           1. .319
                     /gene="IGG2a"
     exon            320. .613
                     /gene="IGG2a"
                     /number=1
     intron          614. .929
                     /gene="IGG2a"
                     /number=1
     CDS             join(<930. .1034,1158. .1487,1571. .1891,3265. .3398,
                     3780. .3863)
                     /gene="IGG2a"
                     /note="alternative splice variant
                     membrane-bound form"
                     /codon_start=3
                     /product="immunoglobulin heavy chain constant region"
                     /protein_id="CAB64865.1"
                     /db_xref="GI:6687182"
                     /translation="PKIPQPQPKPQPQPQPQPKPQPKPEPECTCPKCPAPELLGGPSV
                     FIFPPKPKDVLSISGRPEVTCVVVDVGQEDPEVSFNWYIDGVEVRTANTKPKEEQFNS
                     TYRVVSVLTIQHQDWLTGKELKCKVNNKALPAPIERTISKAKGQTREPQVYTLAPHRE
                     ELAKDTVSITCLVIGFYPADINVEWQRNGRPESEGAYATTLPQLDNDGTYFLYSKLSV
                     GKNTWQQGETFTCVVMHEALHNHSTQKSITQSSEPLLEEESCAEAQSGELDGLWTTIS
                     IFITLFLLSVCYSATVTLFKVKWIFSSVVELKRTIVPDYRNMIGQGS"
```

```
     C_region        join(<930. .1034,1158. .1487,1571. .1891,3265. .3398,
                     3780. .3863)
                     /gene="IGG2a"
                     /product="immunoglobulin heavy chain constant region"
                     /note="alternative splice variant
                     membrane-bound form"
     CDS             join(<930. .1034,1158. .1487,1571. .1899)
                     /gene="IGG2a"
                     /note="alternative splice transcript
                     secreted form"
                     /codon_start=3
                     /product="immunoglobulin heavy chain constant region"
                     /protein_id="CAB64864.1"
                     /db_xref="GI:6687181"
                     /translation="PKIPQPQPKPQPQPQPQPKPQPKPEPECTCPKCPAPELLGGPSV
                     FIFPPKPKDVLSISGRPEVTCVVVDVGQEDPEVSFNWYIDGVEVRTANTKPKEEQFNS
                     TYRVVSVLTIQHQDWLTGKELKCKVNNKALPAPIERTISKAKGQTREPQVYTLAPHRE
                     ELAKDTVSITCLVIGFYPADINVEWQRNGRPESEGAYATTLPQLDNDGTYFLYSKLSV
                     GKNTWQQGETFTCVVMHEALHNHSTQKSITQSSGK"
     C_region        join(<930. .1034,1158. .1487,1571. .1899)
                     /gene="IGG2a"
                     /product="immunoglobulin heavy chain constant region"
                     /note="alternative splice transcript
                     secreted form"
     exon            930. .1034
                     /gene="IGG2a"
                     /number=2
     intron          1035. .1157
                     /gene="IGG2a"
                     /number=2
     exon            1158. .1487
                     /gene="IGG2a"
                     /number=3
     intron          1488. .1570
                     /gene="IGG2a"
                     /number=3
     exon            1571. .1899
                     /gene="IGG2a"
                     /note="number 4a"
     exon            1571. .1891
                     /gene="IGG2a"
                     /note="number 4b"
     intron          1892. .3264
                     /gene="IGG2a"
                     /note="number 4b"
     intron          1897. .3264
                     /gene="IGG2a"
                     /note="number 4a"
     polyA_signal    2000. .2005
                     /gene="IGG2a"
     polyA_site      2033
                     /gene="IGG2a"
     exon            3265. .3398
                     /gene="IGG2a"
                     /note="M1 membrane exon
                     putative"
                     /number=5
     intron          3399. .3779
                     /gene="IGG2a"
                     /number=5
     exon            3780. .3863
                     /gene="IGG2a"
                     /note="M2 membrane exon"
                     /number=6
     polyA_signal    4879. .4884
     polyA_signal    6640. .6645
ORIGIN

Alignment Scores:
Pred. No.:              1.89          Length:         8724
Score:                  76.00         Matches:        11
Percent Similarity:     91.67%        Conservative:   0
Best Local Similarity:  91.67%        Mismatches:     1

Query Match:            90.48%        Indels:         0
DB:                     4             Gaps:           0

US-09-858-332G-15 (1-12) x CDR131945 (1-8724)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||   ||||||||||||||||||||||||||
Db        948 CACAACCAAAACCACAACCACAACCACAACCACAAC 983

RESULT 223
AY191010
LOCUS       AY191010                8867 bp    DNA     linear   PLN 26-FEB-2003
DEFINITION  Zea mays Bz(Dp26)-m1 bz locus tandem duplication sequence.
ACCESSION   AY191010
VERSION     AY191010.1  GI:28569515
KEYWORDS    .
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 8867)
  AUTHORS   Dooner,H.K. and Martinez-Ferez,I.M.
  TITLE     Germinal excisions of the maize transposon activator do not
            stimulate meiotic recombination or homology-dependent repair at the
            bz locus
  JOURNAL   Genetics 147 (4), 1923-1932 (1997)
   PUBMED   9409847
REFERENCE   2  (bases 1 to 8867)
  AUTHORS   Dooner,H.K. and Zheng,Z.
  TITLE     De novo origin of a tandem duplication in maize by improper
            resolution of a recombination intermediate
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 8867)
  AUTHORS   Zheng,Z., Martinez-Ferez,I.M. and Dooner,H.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2002) Waksman Institute, Rutgers University, Hoes
            Lane, Piscataway, NJ 08855, USA
FEATURES             Location/Qualifiers
     source          1. .8867
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="W22"
                     /db_xref="taxon:4577"
                     /chromosome="9"
                     /map="9S; between stk1 and stc1"
                     /note="tandem duplication of bz locus
                     genotype: Bz(Dp26)-m1; Ac-mutable allele"
     misc_feature    1. .6905
                     /note="proximal member of Dp26 duplication; carries an
                     Ac-mutable hybrid bz allele produced by intragenic
                     recombination between the bz-m1 mutant from Bz-McC found
                     in GenBank Accession Numbers AF037598 and X07940 and the
                     bz-E4 mutant from Bz-W22 found in GenBank Accession
                     Numbers AY167674 and X07937"
     repeat_region   1308. .1315
                     /note="target site direct repeat"
     repeat_region   1316. .5880
                     /transposon="Ac"
     repeat_region   5881. .5888
                     /note="target site direct repeat"
     misc_recomb     6906. .6911
                     /note="duplication junction"
     misc_feature    6912. .8867
                     /note="distal member of Dp26 duplication; carries a mutant
                     bz-E4 allele"
     variation       8182
                     /note="E4 mutant site"
ORIGIN

Alignment Scores:
Pred. No.:              1.92          Length:         8867
Score:                  76.00         Matches:        11
```

```
Percent Similarity:     91.67%          Conservative:  0
Best Local Similarity:  91.67%          Mismatches:    1
Query Match:            90.48%          Indels:        0
DB:                     15              Gaps:          0

US-09-858-332G-15 (1-12) x AY191010 (1-8867)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||||||||||||||||||||||   |||||||||
Db       2735 CACAACCACAACCACAACCACAACCAGAACCACAAC 2770

RESULT 224
AY218787/c
LOCUS       AY218787               28708 bp    DNA     circular SYN 21-APR-2003
DEFINITION  Cloning vector pAJ1, complete sequence.
ACCESSION   AY218787 AC138659
VERSION     AY218787.1  GI:30027630
KEYWORDS    .
SOURCE      Cloning vector pAJ1
  ORGANISM  Cloning vector pAJ1
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 28708)
  AUTHORS   McCombie,W.R., Martienssen,R., de la Bastide,M., Preston,R.,
            Spiegel,L., Nascimento,L., Zutavern,T., Balija,V., Bell,M.,
            Miller,B., Muller,S., Katzenberger,F., Yang,C., Dike,S.,
            O'Shaughnessy,A., Dedhia,N. and Palmer,L.
  TITLE     Complete sequence of cloning vector pAJ1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 28708)
  AUTHORS   McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JAN-2003) Lita Annenberg Hazen Genome Center, Cold
            Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor,
            NY 11724, USA
COMMENT     On Apr 21, 2003 this sequence version replaced gi:27753654.
FEATURES             Location/Qualifiers
     source          1. .28708
                     /organism="Cloning vector pAJ1"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:218927"
ORIGIN

Alignment Scores:
Pred. No.:              5.95            Length:        28708
Score:                  76.00           Matches:       11
Percent Similarity:     91.67%          Conservative:  0
Best Local Similarity:  91.67%          Mismatches:    1
Query Match:            90.48%          Indels:        0
DB:                     11              Gaps:          0

US-09-858-332G-15 (1-12) x AY218787 (1-28708)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||||||||||||||||||||||   |||||||||
Db      19313 CACAACCACAACCACAACCACAACCAGAACCACAAC 19278

RESULT 225
AC116305_1/c
WPCOMMENT
  Sequence split into 5 fragments  LOCUS AC116305 Accession AC116305
     Fragment Name      Begin     End
     AC116305_0             1    110000
     AC116305_1        100001    210000
     AC116305_2        200001    310000
     AC116305_3        300001    410000
     AC116305_4        400001    413138
Continuation (2 of 5) of AC116305 from base 100001 (AC116305 Dictyostelium discoideum ch

Alignment Scores:
Pred. No.:              21.7            Length:        110000
Score:                  76.00           Matches:       11
Percent Similarity:     100.00%         Conservative:  0
Best Local Similarity:  100.00%         Mismatches:    0
Query Match:            90.48%          Indels:        0
DB:                     2               Gaps:          0

US-09-858-332G-15 (1-12) x AC116305_1 (1-110000)

Qy          2 AsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||
Db      71545 AACCACAACCACAACCACAACCACAACCATAAC 71513

RESULT 226
BX663527/c
LOCUS       BX663527               111067 bp    DNA     linear   VRT 11-FEB-2004
DEFINITION  Chicken DNA sequence from clone WAG-112A23, complete sequence.
ACCESSION   BX663527
VERSION     BX663527.6  GI:42538816
KEYWORDS    HTG.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 111067)
  AUTHORS   Sycamore,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-FEB-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 11, 2004 this sequence version replaced gi:42406689.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep WAG-112A23 is
            from a Wageningen University White Leghorn chicken 020-CHK-H3 BAC
            library
            VECTOR: pECBAC1.
FEATURES             Location/Qualifiers
     source          1. .111067
                     /organism="Gallus gallus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9031"
                     /clone="WAG-112A23"
                     /clone_lib="WUABG-WL"
ORIGIN

Alignment Scores:
Pred. No.:              21.9            Length:        111067
Score:                  76.00           Matches:       11
Percent Similarity:     91.67%          Conservative:  0
Best Local Similarity:  91.67%          Mismatches:    1
Query Match:            90.48%          Indels:        0
DB:                     5               Gaps:          0
```

```
US-09-858-332G-15 (1-12) x BX663527 (1-111067)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||   ||||||||||||||||||||
Db      31116 CATAACCATAACCAAAACCATAACCATAATCATAAC 31081

RESULT 227
AC004688
LOCUS       AC004688               169794 bp    DNA     linear   HTG 12-AUG-2000
DEFINITION  Plasmodium falciparum chromosome 12 clone 3D7, *** SEQUENCING IN
            PROGRESS ***, 4 unordered pieces.
ACCESSION   AC004688
VERSION     AC004688.8  GI:9797726
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 169794)
  AUTHORS   Hyman,R.W., Fung,E.L., Qin,F., Rowley,D., Mao,J., Tamaki,T.,
            Kurdi,O.B., Conway,A.B. and Davis,R.W.
  TITLE     Plasmodium falciparum 3D7 chromosome 12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169794)
  AUTHORS   Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-1998) Stanford DNA Sequencing and Technology
            Center, Stanford University, 855 California Avenue, Palo Alto, CA
            94304, USA
COMMENT     On Aug 12, 2000 this sequence version replaced gi:8810452.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    43829: contig of 43829 bp in length
            *    43830    44029: gap of unknown length
            *    44030    87202: contig of 43173 bp in length
            *    87203    87402: gap of unknown length
            *    87403    94824: contig of 7422 bp in length
            *    94825    95024: gap of unknown length
            *    95025   169794: contig of 74770 bp in length.
FEATURES             Location/Qualifiers
     source          1. .169794
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5833"
                     /chromosome="12"
                     /clone="PFYAC69"
                     /clone="3D7"
     gap             43830. .44029
                     /estimated_length=unknown
     gap             87203. .87402
                     /estimated_length=unknown
     gap             94825. .95024
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              32.9          Length:         169794
Score:                  76.00         Matches:        11
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            90.48%        Indels:         0
DB:                     14            Gaps:           0

US-09-858-332G-15 (1-12) x AC004688 (1-169794)

Qy          2 AsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||
Db      49926 AATCATAATCATAATCATAATCATAATCATAAT 49958

RESULT 228
AC151411
LOCUS       AC151411               173264 bp    DNA     linear   ROD 28-JAN-2005
DEFINITION  Mus musculus BAC clone RP24-363M23 from 17, complete sequence.
ACCESSION   AC151411
VERSION     AC151411.2  GI:53384961
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 173264)
  AUTHORS   Tomlinson,C., Kozlowicz,A. and Haakenson,W.
  TITLE     The sequence of Mus musculus BAC clone RP24-363M23
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 173264)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-AUG-2004) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 173264)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 173264)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JAN-2005) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Oct 1, 2004 this sequence version replaced gi:51556423.
            ------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            ------------- Summary Statistics
                Center project name: M_BB0363M23
            -------------.

            NOTICE:

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-24 BAC Library has been constructed by Pieter de Jong and
            coworkers (http://www.chori.org) from male C57BL/6J mouse spleen
            and/or brain genomic DNA. The clone and detailed information can be
            obtained from Pieter de Jong and coworkers at http://www.chori.org

            NEIGHBORING SEQUENCE INFORMATION:
            This sequence is the entire insert of the clone.
FEATURES             Location/Qualifiers
     source          1. .173264
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="17"
```

```
                     /map="17"
                     /clone="RP24-363M23"
                     /clone_lib="RPCI-24"
     unsure          72488. .74098
                     /note="Unresolved tandem repeat."
     unsure          112614. .112807
                     /note="Unresolved simple sequence repeat."
     repeat_region   168814. .168815
                     /transposon="Bacterial transposon insertion in clone
                     excised here."
ORIGIN

Alignment Scores:
Pred. No.:                33.6          Length:          173264
Score:                    76.00         Matches:         11
Percent Similarity:       91.67%        Conservative:    0
Best Local Similarity:    91.67%        Mismatches:      1
Query Match:              90.48%        Indels:          0
DB:                       9             Gaps:            0

US-09-858-332G-15 (1-12) x AC151411 (1-173264)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||   ||||||||||||||||||||||
Db      15737 CACAACCACAACCAAAACCACAACCACAACCACAAC 15772

RESULT 229
AC097154/c
LOCUS       AC097154               177414 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-217G16, *** SEQUENCING IN PROGRESS
            ***.
ACCESSION   AC097154
VERSION     AC097154.6  GI:30521144
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 177414)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 177414)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 177414)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:22855236.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GIES
                Center clone name: CH230-217G16
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 169050 bases at least Q40
                Consensus quality: 170260 bases at least Q30
                Consensus quality: 171096 bases at least Q20
                Estimated insert size: 175309; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   177414: contig of 177414 bp in length.
FEATURES                Location/Qualifiers
     source             1. .177414
```

```
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-217G16"
     misc_feature    1. .1155
                     /note="wgs_contig"
     misc_feature    173483. .175537
                     /note="wgs_contig"
     misc_feature    175588. .177414
                     /note="wgs_contig"
ORIGIN

Alignment Scores:
Pred. No.:                 34.4          Length:          177414
Score:                     76.00         Matches:         11
Percent Similarity:        100.00%       Conservative:    0
Best Local Similarity:     100.00%       Mismatches:      0
Query Match:               90.48%        Indels:          0
DB:                        14            Gaps:            0

US-09-858-332G-15 (1-12) x AC097154 (1-177414)

Qy          2 AsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||
Db       9727 AATCACAATCATAATCACAATCATAATCATAAT 9695

RESULT 230
AC068786
LOCUS       AC068786               178733 bp    DNA     linear   HTG 08-MAR-2001
DEFINITION  Homo sapiens chromosome 19 clone RP11-50I11, WORKING DRAFT
            SEQUENCE, 15 unordered pieces.
ACCESSION   AC068786
VERSION     AC068786.11  GI:13249431
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 178733)
  AUTHORS   Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
            Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
            Mao,J., Lam,B., Marathe,R., Miranda,M., Morehouse,A.J., Nguyen,M.,
            Oefner,P., Palm,C.J., Ramirez,D., Southwick,A.M., Wilhelmy,J.,
            Yu,S. and Davis,R.W.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 178733)
  AUTHORS   Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
            Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
            Mao,J., Marathe,R., Morehouse,A.J., Oefner,P., Palm,C.J.,
            Ramirez,D., Wilhelmy,J., Yu,S. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2000) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
COMMENT     On Mar 8, 2001 this sequence version replaced gi:13236646.
            -------------- Genome Center
                 Center: Stanford DNA Sequencing and Technology Development
            Center
                 Center code: SDSTDC
                 Web site: http://sequence-www.stanford.edu/group/human/
                 Contact: hum-info@sequence.stanford.edu
            -------------- Project Information
                 Center project name: 899
                 Center clone name: RP11-50I11
            -------------- Summary Statistics
                 Sequencing Vector: M13mp18; X02513
                 Chemistry: Dye-primer; 0% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 168546 bases at least Q40
                 Consensus quality: 171030 bases at least Q30
                 Consensus quality: 172366 bases at least Q20
                 Insert size: 178045; agarose-fp
                 Insert size: 177333; sum-of-contigs
                 Quality coverage:  9.2x in Q20 bases; agarose-fp
                 Quality coverage:  9.2x in Q20 bases; sum-of-contigs.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 15 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1625: contig of 1625 bp in length
            *     1626     1725: gap of unknown length
            *     1726     3491: contig of 1766 bp in length
            *     3492     3591: gap of unknown length
            *     3592     5210: contig of 1619 bp in length
            *     5211     5310: gap of unknown length
            *     5311     7665: contig of 2355 bp in length
            *     7666     7765: gap of unknown length
            *     7766    11506: contig of 3741 bp in length
            *    11507    11606: gap of unknown length
            *    11607    16778: contig of 5172 bp in length
            *    16779    16878: gap of unknown length
            *    16879    24169: contig of 7291 bp in length
            *    24170    24269: gap of unknown length
            *    24270    34333: contig of 10064 bp in length
            *    34334    34433: gap of unknown length
            *    34434    46447: contig of 12014 bp in length
            *    46448    46547: gap of unknown length
            *    46548    58761: contig of 12214 bp in length
            *    58762    58861: gap of unknown length
            *    58862    73753: contig of 14892 bp in length
            *    73754    73853: gap of unknown length
            *    73854    92184: contig of 18331 bp in length
            *    92185    92284: gap of unknown length
            *    92285   114445: contig of 22161 bp in length
            *   114446   114545: gap of unknown length
            *   114546   140410: contig of 25865 bp in length
            *   140411   140510: gap of unknown length
            *   140511   178733: contig of 38223 bp in length.
FEATURES             Location/Qualifiers
     source          1. .178733
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="RP11-50I11"
                     /clone_lib="RPCI human BAC library 11"
     misc_feature    1. .1625
                     /note="assembly_name:Contig21"
     gap             1626. .1725
                     /estimated_length=unknown
     misc_feature    1726. .3491
                     /note="assembly_name:Contig24"
     gap             3492. .3591
                     /estimated_length=unknown
     misc_feature    3592. .5210
                     /note="assembly_name:Contig30"
     gap             5211. .5310
                     /estimated_length=unknown
     misc_feature    5311. .7665
                     /note="assembly_name:Contig32"
     gap             7666. .7765
                     /estimated_length=unknown
     misc_feature    7766. .11506
                     /note="assembly_name:Contig33"
     gap             11507. .11606
                     /estimated_length=unknown
     misc_feature    11607. .16778
                     /note="assembly_name:Contig34"
     gap             16779. .16878
                     /estimated_length=unknown
```

```
     misc_feature    16879. .24169
                     /note="assembly_name:Contig35"
     gap             24170. .24269
                     /estimated_length=unknown
     misc_feature    24270. .34333
                     /note="assembly_name:Contig36"
     gap             34334. .34433
                     /estimated_length=unknown
     misc_feature    34434. .46447
                     /note="assembly_name:Contig37"
     gap             46448. .46547
                     /estimated_length=unknown
     misc_feature    46548. .58761
                     /note="assembly_name:Contig38
                     clone_end:T7"
     gap             58762. .58861
                     /estimated_length=unknown
     misc_feature    58862. .73753
                     /note="assembly_name:Contig39"
     gap             73754. .73853
                     /estimated_length=unknown
     misc_feature    73854. .92184
                     /note="assembly_name:Contig40"
     gap             92185. .92284
                     /estimated_length=unknown
     misc_feature    92285. .114445
                     /note="assembly_name:Contig41
                     clone_end:SP6"
     gap             114446. .114545
                     /estimated_length=unknown
     misc_feature    114546. .140410
                     /note="assembly_name:Contig42"
     gap             140411. .140510
                     /estimated_length=unknown
     misc_feature    140511. .178733
                     /note="assembly_name:Contig43"
ORIGIN

Alignment Scores:
Pred. No.:              34.6          Length:        178733
Score:                  76.00         Matches:       11
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            90.48%        Indels:        0
DB:                     14            Gaps:          0

US-09-858-332G-15 (1-12) x AC068786 (1-178733)

Qy          2 AsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||
Db      37203 AATCATAATCATAATCATAATCATAATCATAAT 37235

RESULT 231
AC121770/c
LOCUS       AC121770              186190 bp    DNA     linear   ROD 27-JAN-2005
DEFINITION  Mus musculus BAC clone RP23-341M16 from 17, complete sequence.
ACCESSION   AC121770
VERSION     AC121770.3  GI:55734069
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 186190)
  AUTHORS   Wan-Ching,W.-ching. and Cotton,M.
  TITLE     The sequence of Mus musculus BAC clone RP23-341M16
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 186190)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAY-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 186190)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-2004) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 186190)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-2004) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   5  (bases 1 to 186190)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2005) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Nov 13, 2004 this sequence version replaced gi:54400883.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: M_BA0341M16
            --------------.

            NOTICE:

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-23 BAC Library has been constructed by Kazutoyo Osegawa
            and Minako Tateno in the laboratory of Pieter de Jong
            (http://www.chori.org) from female C57BL/6J mouse kidney and/or
            brain genomic DNA.  The clone and detailed information can be
            obtained from Research Genetics, Inc. (http://www.resgen.com) or
            Pieter de Jong and coworkers at http://www.chori.org

            NEIGHBORING SEQUENCE INFORMATION:
            This sequence is the entire insert of the clone.
FEATURES             Location/Qualifiers
     source          1. .186190
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="17"
                     /map="17"
                     /clone="RP23-341M16"
                     /clone_lib="RPCI-23"
     unsure          44692. .44959
                     /note="Unresolved simple sequence repeat."
     unsure          44728. .44801
                     /note="Sequence derived from one plasmid subclone."
ORIGIN

Alignment Scores:
Pred. No.:              36            Length:        186190
Score:                  76.00         Matches:       11
Percent Similarity:     91.67%        Conservative:  0
Best Local Similarity:  91.67%        Mismatches:    1
Query Match:            90.48%        Indels:        0
```

```
DB:                    9             Gaps:          0

US-09-858-332G-15 (1-12) x AC121770 (1-186190)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||   ||||||||||||||||||||
Db     141557 CACAACCACAACCAAAACCACAACCACAACCACAAC 141522

RESULT 232
AC147567/c
LOCUS       AC147567               191770 bp    DNA     linear   ROD 27-JAN-2005
DEFINITION  Mus musculus BAC clone RP23-107E14 from 6, complete sequence.
ACCESSION   AC147567
VERSION     AC147567.3  GI:51921474
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 191770)
  AUTHORS   Kruchowski,S., Kozlowicz,A. and Meyer,R.
  TITLE     The sequence of Mus musculus BAC clone RP23-107E14
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 191770)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 191770)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-AUG-2004) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 191770)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2004) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   5  (bases 1 to 191770)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2005) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Sep 8, 2004 this sequence version replaced gi:51241981.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: M_BA0107E14
            --------------.

            NOTICE:

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-23 BAC Library has been constructed by Kazutoyo Osegawa
            and Minako Tateno in the laboratory of Pieter de Jong
            (http://www.chori.org) from female C57BL/6J mouse kidney and/or
            brain genomic DNA.  The clone and detailed information can be
            obtained from Research Genetics, Inc. (http://www.resgen.com) or
            Pieter de Jong and coworkers at http://www.chori.org

            NEIGHBORING SEQUENCE INFORMATION:
            This sequence is the entire insert of the clone.
FEATURES                Location/Qualifiers
     source             1. .191770
                        /organism="Mus musculus"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:10090"
                        /chromosome="6"
                        /map="6"
                        /clone="RP23-107E14"
                        /clone_lib="RPCI-23"
     repeat_region      1585. .1704
                        /rpt_family="MaLR"
     repeat_region      1828. .2044
                        /rpt_family="MaLR"
     repeat_region      2240. .2316
                        /rpt_family="ERVK"
     repeat_region      3769. .3833
                        /rpt_family="MIR"
     repeat_region      4949. .5371
                        /rpt_family="MaLR"
     repeat_region      6292. .6402
                        /rpt_family="B4"
     repeat_region      6336. .6597
                        /rpt_family="RMER15"
     repeat_region      7062. .7279
                        /rpt_family="L1"
     repeat_region      7317. .7388
                        /rpt_family="MER1_type"
     repeat_region      7471. .7665
                        /rpt_family="L1"
     repeat_region      7700. .8054
                        /rpt_family="L1"
     repeat_region      10230. .10273
                        /rpt_family="B4"
     repeat_region      10274. .10346
                        /rpt_family="ID"
     repeat_region      10568. .10777
                        /rpt_family="MER2_type"
     repeat_region      11108. .11176
                        /rpt_family="B4"
     repeat_region      11177. .12558
                        /rpt_family="L1"
     repeat_region      13162. .13307
                        /rpt_family="L2"
     repeat_region      13600. .13834
                        /rpt_family="MaLR"
     repeat_region      14010. .14284
                        /rpt_family="B4"
     repeat_region      14819. .15148
                        /rpt_family="L1"
     repeat_region      15339. .15530
                        /rpt_family="L1"
     repeat_region      15531. .15754
                        /rpt_family="B2"
     repeat_region      15755. .16034
                        /rpt_family="L1"
     repeat_region      16263. .16477
                        /rpt_family="MER1_type"
     repeat_region      16520. .17575
                        /rpt_family="L1"
     repeat_region      18529. .18831
                        /rpt_family="B4"
     repeat_region      20442. .20568
                        /rpt_family="L1"
     repeat_region      22229. .22301
                        /rpt_family="ID"
```

```
     repeat_region   22340. .22534
                     /rpt_family="MIR"
     repeat_region   22771. .22858
                     /rpt_family="Alu"
     repeat_region   24560. .25239
                     /rpt_family="L1"
     repeat_region   25289. .25984
                     /rpt_family="L1"
     unsure          26156. .26477
                     /note="Sequence derived from one plasmid subclone."
     repeat_region   26860. .26930
                     /rpt_family="MIR"
     repeat_region   28026. .28109
                     /rpt_family="Alu"
     repeat_region   28110. .28122
                     /rpt_family="B4"
     repeat_region   29919. .30354
                     /rpt_family="L1"
     repeat_region   31286. .31378
                     /rpt_family="Alu"
     repeat_region   31440. .31684
                     /rpt_family="MER2_type"
     repeat_region   32247. .32405
                     /rpt_family="B4"
     repeat_region   32683. .33014
                     /rpt_family="MaLR"
     repeat_region   33513. .33662
                     /rpt_family="B4"
     repeat_region   33663. .33711
                     /rpt_family="ERV1"
     repeat_region   35869. .36123
                     /rpt_family="B4"
     repeat_region   36812. .36992
                     /rpt_family="ERVK"
     repeat_region   38843. .38977
                     /rpt_family="B4"
     repeat_region   38978. .39019
                     /rpt_family="B2"
     repeat_region   40303. .40654
                     /rpt_family="MaLR"
     repeat_region   40801. .40968
                     /rpt_family="ERVK"
     repeat_region   42055. .42264
                     /rpt_family="B2"
     repeat_region   45105. .45283
                     /rpt_family="MER2_type"
     repeat_region   46743. .46913
                     /rpt_family="MIR"
     repeat_region   49728. .50276
                     /rpt_family="L1"
     repeat_region   51262. .51306
                     /rpt_family="MIR"
     repeat_region   52301. .52375
                     /rpt_family="MIR"
     repeat_region   52681. .52820
                     /rpt_family="MIR"
     repeat_region   54545. .54726
                     /rpt_family="MER1_type"
     repeat_region   55031. .55188
                     /rpt_family="MER1_type"
     repeat_region   56471. .56542
                     /rpt_family="MER1_type"
     repeat_region   59108. .59255
                     /rpt_family="B4"

Alignment Scores:
Pred. No.:              37              Length:         191770
Score:                  76.00           Matches:        11
Percent Similarity:     91.67%          Conservative:   0
Best Local Similarity:  91.67%          Mismatches:     1
Query Match:            90.48%          Indels:         0
DB:                     9               Gaps:           0
```

```
US-09-858-332G-15 (1-12) x AC147567 (1-191770)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||||||||||||||||   ||||||||||||||||
Db     144196 CACAACCACAACCACAACCAAAACCACAATCACAAC 144161

RESULT 233
AC004709
LOCUS       AC004709              196149 bp    DNA     linear   HTG 01-APR-1999
DEFINITION  Plasmodium falciparum chromosome 12, *** SEQUENCING IN PROGRESS
            ***, 3 unordered pieces.
ACCESSION   AC004709
VERSION     AC004709.3  GI:4558585
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 196149)
  AUTHORS   Hyman,R.W., Fung,E.L., Qin,F., Tamaki,T., Kurdi,O.B., Conway,A.B.
            and Davis,R.W.
  TITLE     Plasmodium falciparum 3D7 chromosome 12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 196149)
  AUTHORS   Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAY-1998) Stanford DNA Sequencing and Technology
            Center, Stanford University, 855 California Avenue, Palo Alto, CA
            94304, USA
COMMENT     On Apr 2, 1999 this sequence version replaced gi:4337173.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    47631: contig of 47631 bp in length
            *    47632    47831: gap of unknown length
            *    47832   179129: contig of 131298 bp in length
            *   179130   179329: gap of unknown length
            *   179330   196149: contig of 16820 bp in length.
FEATURES             Location/Qualifiers
     source          1. .196149
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5833"
                     /chromosome="12"
     gap             47632. .47831
                     /estimated_length=unknown
     gap             179130. .179329
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              37.8            Length:         196149
Score:                  76.00           Matches:        11
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            90.48%          Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-15 (1-12) x AC004709 (1-196149)

Qy          2 AsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||
Db     141271 AATCATAATCATAATCATAATCATAATCATAAT 141303

RESULT 234
AC140676
LOCUS       AC140676              197819 bp    DNA     linear   ROD 27-JAN-2005
DEFINITION  Mus musculus BAC clone RP23-124A12 from 6, complete sequence.
ACCESSION   AC140676
```

```
VERSION     AC140676.3  GI:52078001
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 197819)
  AUTHORS   Hlafka,M., Cotton,M. and Bielicki,L.
  TITLE     The sequence of Mus musculus BAC clone RP23-124A12
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 197819)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-FEB-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 197819)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-FEB-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 197819)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-2004) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   5  (bases 1 to 197819)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2005) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Sep 15, 2004 this sequence version replaced gi:28570546.
            ------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            ------------- Summary Statistics
                Center project name: M_BA0124A12
            -------------.

            NOTICE:

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-23 BAC Library has been constructed by Kazutoyo Osegawa
            and Minako Tateno in the laboratory of Pieter de Jong
            (http://www.chori.org) from female C57BL/6J mouse kidney and/or
            brain genomic DNA.  The clone and detailed information can be
            obtained from Research Genetics, Inc. (http://www.resgen.com) or
            Pieter de Jong and coworkers at http://www.chori.org

            NEIGHBORING SEQUENCE INFORMATION:
            This sequence is the entire insert of the clone.
FEATURES            Location/Qualifiers
     source         1. .197819
                    /organism="Mus musculus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10090"
                    /chromosome="6"
                    /map="6"
                    /clone="RP23-124A12"
                    /clone_lib="RPCI-23"
     repeat_region  925. .1044
                    /rpt_family="Alu"
     repeat_region  1220. .1573
                    /rpt_family="MaLR"
     repeat_region  2693. .2829
                    /rpt_family="L1"
     repeat_region  2874. .3181
                    /rpt_family="L1"
     repeat_region  3434. .3736
                    /rpt_family="B4"
     repeat_region  4715. .5000
                    /rpt_family="B4"
     repeat_region  5336. .5514
                    /rpt_family="MIR"
     repeat_region  6453. .6500
                    /rpt_family="ERV1"
     repeat_region  7294. .7352
                    /rpt_family="ID"
     repeat_region  7372. .7479
                    /rpt_family="L1"
     repeat_region  7533. .7824
                    /rpt_family="L1"
     repeat_region  8334. .8538
                    /rpt_family="B2"
     repeat_region  9300. .9507
                    /rpt_family="B4"
     repeat_region  11437. .11514
                    /rpt_family="ID"
     repeat_region  13163. .13500
                    /rpt_family="L1"
     repeat_region  14720. .14866
                    /rpt_family="B4"
     repeat_region  15236. .15297
                    /rpt_family="Tip100"
     repeat_region  16373. .16749
                    /rpt_family="MaLR"
     repeat_region  17078. .17151
                    /rpt_family="L1"
     repeat_region  18435. .18581
                    /rpt_family="Alu"
     repeat_region  24021. .24293
                    /rpt_family="MaLR"
     repeat_region  24294. .24482
                    /rpt_family="B2"
     repeat_region  24483. .24544
                    /rpt_family="MaLR"
     repeat_region  24750. .24877
                    /rpt_family="MIR"
     repeat_region  24880. .24959
                    /rpt_family="MIR"
     repeat_region  26334. .26607
                    /rpt_family="B4"
     repeat_region  27131. .27268
                    /rpt_family="B4"
     repeat_region  28629. .28696
                    /rpt_family="ID"
     repeat_region  29062. .29399
                    /rpt_family="MaLR"
     repeat_region  30170. .30216
                    /rpt_family="B4"
     repeat_region  30624. .30845
                    /rpt_family="B4"
     repeat_region  31611. .31689
                    /rpt_family="MIR"
     repeat_region  31898. .32333
                    /rpt_family="L1"
     repeat_region  33584. .33638
                    /rpt_family="ERVK"
     repeat_region  34243. .34331
```

```
                     /rpt_family="MaLR"
     repeat_region   34332. .34473
                     /rpt_family="L1"
     repeat_region   34483. .34791
                     /rpt_family="MaLR"
     repeat_region   35108. .35479
                     /rpt_family="L1"
     repeat_region   36020. .36121
                     /rpt_family="MIR"
     repeat_region   37665. .37986
                     /rpt_family="L2"
     repeat_region   38305. .38434
                     /rpt_family="L1"
     repeat_region   39500. .39705
                     /rpt_family="B2"
     repeat_region   39565. .39707
                     /rpt_family="B2"
     repeat_region   42036. .42140
                     /rpt_family="L1"
     repeat_region   44032. .44102
                     /rpt_family="MIR"
     repeat_region   44725. .44872
                     /rpt_family="B4"
     repeat_region   45650. .45839
                     /rpt_family="MaLR"
     repeat_region   46011. .46200
                     /rpt_family="MaLR"
     repeat_region   48353. .48494
                     /rpt_family="MER1_type"
     repeat_region   48876. .49098
                     /rpt_family="MER2_type"
     repeat_region   49695. .50056
                     /rpt_family="MaLR"
     repeat_region   50706. .50774
                     /rpt_family="ERV1"
     repeat_region   50951. .51005
                     /rpt_family="L1"
     repeat_region   52132. .52320
                     /rpt_family="L2"
     unsure          52687. .52894
                     /note="Unresolved simple sequence repeat."
     repeat_region   52921. .53031
                     /rpt_family="B4"
     repeat_region   55577. .55680
                     /rpt_family="Alu"
     repeat_region   61177. .61298
                     /rpt_family="B4"
     repeat_region   61400. .61609
                     /rpt_family="B2"
     repeat_region   62099. .62215
                     /rpt_family="BC1_MM"

Alignment Scores:
Pred. No.:              38.2          Length:         197819
Score:                  76.00         Matches:        11
Percent Similarity:     91.67%        Conservative:   0
Best Local Similarity:  91.67%        Mismatches:     1
Query Match:            90.48%        Indels:         0
DB:                     9             Gaps:           0

US-09-858-332G-15 (1-12) x AC140676 (1-197819)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||   ||||||||||||||
Db      92801 CACAACCACAACCACAACCAAAACCACAATCACAAC 92836

RESULT 235
AC123264/c
LOCUS       AC123264              215675 bp    DNA     linear   HTG 08-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-290F24, *** SEQUENCING IN PROGRESS
            ***, 4 unordered pieces.
ACCESSION   AC123264
VERSION     AC123264.3  GI:22856538
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 215675)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 215675)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 215675)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 14, 2002 this sequence version replaced gi:21671638.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
```

(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GTUK
Center clone name: CH230-290F24
---------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 199476 bases at least Q40
Consensus quality: 202697 bases at least Q30
Consensus quality: 204497 bases at least Q20
Estimated insert size: 220130; sum-of-contigs estimation
Quality coverage: 3x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 209764: contig of 209764 bp in length
* 209765 209864: gap of unknown length
* 209865 210963: contig of 1099 bp in length
* 210964 211063: gap of unknown length
* 211064 212266: contig of 1203 bp in length
* 212267 212366: gap of unknown length
* 212367 215675: contig of 3309 bp in length.

FEATURES Location/Qualifiers
source 1. .215675
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-290F24"
misc_feature 1. .1274
/note="wgs_contig"
gap 209765. .209864
/estimated_length=unknown
gap 210964. .211063
/estimated_length=unknown
gap 212267. .212366
/estimated_length=unknown

ORIGIN

Alignment Scores:

Pred. No.:	41.5	Length:	215675
Score:	76.00	Matches:	11
Percent Similarity:	91.67%	Conservative:	0
Best Local Similarity:	91.67%	Mismatches:	1
Query Match:	90.48%	Indels:	0
DB:	14	Gaps:	0

US-09-858-332G-15 (1-12) x AC123264 (1-215675)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||   ||||||||||||||||||||||
Db     167889 CACAACCACAACCAGAACCACAACCACAACCACAAC 167854
```

RESULT 236
AC157594

LOCUS	AC157594 220526 bp DNA linear ROD 21-JUN-2005
DEFINITION	Mus musculus BAC clone RP23-279A2 from chromosome 17, complete sequence.
ACCESSION	AC157594
VERSION	AC157594.4 GI:66865188
KEYWORDS	HTG.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 220526)
AUTHORS	Harkins,R., Haglund,K. and Bielicki,L.
TITLE	The sequence of Mus musculus BAC clone RP23-279A2
JOURNAL	Unpublished (2001)
REFERENCE	2 (bases 1 to 220526)
AUTHORS	Wilson,R.K.
TITLE	Direct Submission
JOURNAL	Submitted (22-FEB-2005) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE	3 (bases 1 to 220526)
AUTHORS	Wilson,R.K.
TITLE	Direct Submission
JOURNAL	Submitted (06-MAY-2005) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE	4 (bases 1 to 220526)
AUTHORS	Wilson,R.K.
TITLE	Direct Submission
JOURNAL	Submitted (02-JUN-2005) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE	5 (bases 1 to 220526)
AUTHORS	Wilson,R.K.
TITLE	Direct Submission
JOURNAL	Submitted (21-JUN-2005) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT	On Jun 2, 2005 this sequence version replaced gi:63055265.

---------- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
---------- Summary Statistics
Center project name: M_BA0279A02

NOTICE:

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e. phred quality >=30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone, fosmid clone or direct clone walk sequence. Sequence from the Mouse Genome Sequencing Consortium whole genome shotgun may have been used to obtain the consensus sequence. The assembly was confirmed by restriction digest.
This finishing standard has slightly changed from the previous Human standard. Specifically, standards for regions of low sequence complexity (such as dinucleotide repeats and small unit tandem repeats) have been relaxed. These regions are very prevalent in the mouse genome, and the return on extended finishing efforts is minimal.
If a sequence meets the criteria of the above statement, it needs no comments or tags. If the criteria are not met, such as ambiguous bases, then the region is duly annotated.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see

```
          http://genome.wustl.edu

          SOURCE INFORMATION:
          The  BAC Library has been constructed by Kazutoyo Osegawa and
          Minako Tateno in the laboratory of Pieter de Jong
          (http://www.chori.org) from female C57BL/6J mouse kidney and/or
          brain genomic DNA.  The clone and detailed information can be
          obtained from Research Genetics, Inc. (http://www.resgen.com) or
          Pieter de Jong and coworkers at http://www.chori.org

          This sequence is the entire insert of the clone.
FEATURES             Location/Qualifiers
     source          1. .220526
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="17"
                     /clone="RP23-279A2"
                     /clone_lib="RPCI-23"
     misc_feature    1. .126
                     /note="Sequence derived from PCR product of project DNA"
     unsure          807. .837
                     /note="Sequence derived from one plasmid subclone."
     unsure          35509. .35537
                     /note="Sequence derived from one plasmid subclone."
     unsure          185314. .185581
                     /note="Unresolved simple sequence repeat."
ORIGIN

Alignment Scores:
Pred. No.:              42.4          Length:        220526
Score:                  76.00         Matches:       11
Percent Similarity:     91.67%        Conservative:  0
Best Local Similarity:  91.67%        Mismatches:    1
Query Match:            90.48%        Indels:        0
DB:                     9             Gaps:          0

US-09-858-332G-15 (1-12) x AC157594 (1-220526)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||   ||||||||||||||||||||
Db      88717 CACAACCACAACCAAAACCACAACCACAACCACAAC 88752

RESULT 237
AC165427/c
LOCUS       AC165427              232340 bp    DNA     linear   HTG 13-JUL-2005
DEFINITION  Mus musculus chromosome 17 clone RP23-55P24, WORKING DRAFT
            SEQUENCE, 6 unordered pieces.
ACCESSION   AC165427
VERSION     AC165427.1  GI:70778720
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 232340)
  AUTHORS   Wilson,R.K.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 232340)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2005) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT
            ----------------- Genome Center ---------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu
            Contact: submissions@watson.wustl.edu
            -------------- Project Information -------------------
            Center project name: M_BA0055P24
            ------------------------------------------------------
            -------------- Summary Statistics  -------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 228923 bases at least Q40
            Consensus quality: 229252 bases at least Q30
            Consensus quality: 229550 bases at least Q20
            ------------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1609: contig of 1609 bp in length
            *     1610     1709: gap of unknown length
            *     1710     3172: contig of 1463 bp in length
            *     3173     3272: gap of unknown length
            *     3273     4793: contig of 1521 bp in length
            *     4794     4893: gap of unknown length
            *     4894    66540: contig of 61647 bp in length
            *    66541    66640: gap of unknown length
            *    66641   134433: contig of 67793 bp in length
            *   134434   134533: gap of unknown length
            *   134534   232340: contig of 97807 bp in length.
FEATURES             Location/Qualifiers
     source          1. .232340
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="17"
                     /clone="RP23-55P24"
     misc_feature    1. .1609
                     /note="assembly_name:Contig10"
     gap             1610. .1709
                     /estimated_length=unknown
     misc_feature    1710. .3172
                     /note="assembly_name:Contig11"
     gap             3173. .3272
                     /estimated_length=unknown
     misc_feature    3273. .4793
                     /note="assembly_name:Contig13
                     clone_end:T7
                     vector_side:left"
     gap             4794. .4893
                     /estimated_length=unknown
     misc_feature    4894. .66540
                     /note="assembly_name:Contig14"
     gap             66541. .66640
                     /estimated_length=unknown
     misc_feature    66641. .134433
                     /note="assembly_name:Contig15"
     gap             134434. .134533
                     /estimated_length=unknown
     misc_feature    134534. .232340
                     /note="assembly_name:Contig16"
ORIGIN

Alignment Scores:
Pred. No.:              44.5          Length:        232340
Score:                  76.00         Matches:       11
Percent Similarity:     91.67%        Conservative:  0
Best Local Similarity:  91.67%        Mismatches:    1
Query Match:            90.48%        Indels:        0
DB:                     14            Gaps:          0

US-09-858-332G-15 (1-12) x AC165427 (1-232340)
```

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||  ||||||||||||||||||||
Db     142207 CACAACCACAACCAAAACCACAACCACAACCACAAC 142172

RESULT 238
AC095709
LOCUS       AC095709               241828 bp    DNA     linear   HTG 09-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-5C3, WORKING DRAFT SEQUENCE.
ACCESSION   AC095709
VERSION     AC095709.6  GI:24817923
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 241828)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,C.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 241828)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 241828)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 9, 2002 this sequence version replaced gi:22773052.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GDCJ
                Center clone name: CH230-5C3
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 234755 bases at least Q40
                Consensus quality: 236298 bases at least Q30
                Consensus quality: 237126 bases at least Q20
                Estimated insert size: 242830; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   241828: contig of 241828 bp in length.
FEATURES                Location/Qualifiers
     source             1. .241828
                        /organism="Rattus norvegicus"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:10116"
                        /clone="CH230-5C3"
     misc_feature       1. .1366
                        /note="wgs_end_extension
                        clone_end:T7"
     misc_feature       complement(5782. .6544)
                        /note="clone_boundary
                        clone_end:T7
                        site:EcoRI
                        end_sequence:BH334709"
     misc_feature       complement(237566. .238481)
                        /note="clone_boundary
                        clone_end:Sp6
                        site:EcoRI
                        end_sequence:BH334711"
     misc_feature       239002. .241828
                        /note="wgs_end_extension
                        clone_end:Sp6"
ORIGIN

Alignment Scores:
```

```
Pred. No.:              46.3          Length:        241828
Score:                  76.00         Matches:       11
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            90.48%        Indels:        0
DB:                     14            Gaps:          0

US-09-858-332G-15 (1-12) x AC095709 (1-241828)

Qy          2 AsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||
Db     192695 AACCACAACCACAACCACAACCACAACCACAAC 192727

RESULT 239
AE014847
LOCUS       AE014847               252650 bp    DNA     linear   INV 11-FEB-2003
DEFINITION  Plasmodium falciparum 3D7 chromosome 12, section 4 of 9 of the
            complete sequence.
ACCESSION   AE014847 AE014188
VERSION     AE014847.1  GI:23496718
KEYWORDS    .
SOURCE      Plasmodium falciparum 3D7
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 252650)
  AUTHORS   Gardner,M.J., Hall,N., Fung,E., White,O., Berriman,M., Hyman,R.W.,
            Carlton,J.M., Pain,A., Nelson,K.E., Bowman,S., Paulsen,I.T.,
            James,K., Eisen,J.A., Rutherford,K., Salzberg,S.L., Craig,A.,
            Kyes,S., Chan,M.S., Nene,V., Shallom,S.J., Suh,B., Peterson,J.,
            Angiuoli,S., Pertea,M., Allen,J., Selengut,J., Haft,D.,
            Mather,M.W., Vaidya,A.B., Martin,D.M., Fairlamb,A.H.,
            Fraunholz,M.J., Roos,D.S., Ralph,S.A., McFadden,G.I.,
            Cummings,L.M., Subramanian,G.M., Mungall,C., Venter,J.C.,
            Carucci,D.J., Hoffman,S.L., Newbold,C., Davis,R.W., Fraser,C.M. and
            Barrell,B.
  TITLE     Genome sequence of the human malaria parasite Plasmodium falciparum
  JOURNAL   Nature 419 (6906), 498-511 (2002)
   PUBMED   12368864
REFERENCE   2  (bases 1 to 252650)
  AUTHORS   Hyman,R.W., Fung,E., Conway,A., Kurdi,O., Mao,J., Miranda,M.,
            Nakao,B., Rowley,D., Tamaki,T., Wang,F. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-2002) Stanford Genome Technology Center, Stanford
            University, 855 California Avenue, Palo Alto, CA 94304, USA
REFERENCE   3  (bases 1 to 252650)
  AUTHORS   Hyman,R.W., Fung,E., Conway,A., Kurdi,O., Mao,J., Miranda,M.,
            Nakao,B., Rowley,D., Tamaki,T., Wang,F. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-2003) Stanford Genome Technology Center, Stanford
            University, 855 California Avenue, Palo Alto, CA 94304, USA
FEATURES             Location/Qualifiers
     source          1. .252650
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="genomic DNA"
                     /isolate="3D7"
                     /db_xref="taxon:36329"
                     /chromosome="12"
     repeat_region   90. .134
                     /rpt_type=tandem
     repeat_region   96. .134
                     /rpt_type=tandem
     repeat_region   586. .669
                     /rpt_type=tandem
     repeat_region   724. .748
                     /rpt_type=tandem
     repeat_region   1220. .1250
                     /rpt_type=tandem
     repeat_region   1331. .1363
                     /rpt_type=tandem
     repeat_region   1440. .1476
                     /rpt_type=tandem
     repeat_region   1758. .1801
                     /rpt_type=tandem
     repeat_region   1907. .1971
                     /rpt_type=tandem
     repeat_region   2150. .2184
                     /rpt_type=tandem
     repeat_region   2301. .2344
                     /rpt_type=tandem
     repeat_region   2484. .2513
                     /rpt_type=tandem
     gene            <2608. .>8601
                     /locus_tag="PFL0930w"
     mRNA            <2608. .>8601
                     /locus_tag="PFL0930w"
     CDS             2608. .8601
                     /locus_tag="PFL0930w"
                     /note="HMMPfam hit to PF00637, 7-fold repeat in Clathrin
                     and VPS, score 2.1e-24"
                     /codon_start=1
                     /product="clathrin heavy chain, putative"
                     /protein_id="AAN36274.1"
                     /db_xref="GI:23496719"
                     /translation="MSQNNPLSVCVADNLINYDIQNESFRLGNVSVEGDKYICVKENV
                     NENTQVVVINLHNKNSTRKHMKAESVIIHPNDPILALKGTIKNMNTIFLQVFNIETKE
                     KICSLNLNEYMNYWKWINNDTIAIVCEKNVYHWNIDIHNTKKNKDNNDNNNNNNNNNN
                     NNNNNNYFNTLTKVFEKAQIFIDNNSQILYYSTDKDMKWCILCGISTQDQGKSIDGYM
                     QLYSCEKKLHQTIEGFIGCFGSFVFDNLDMKPLFCFVEKKKNSTISRLHLMDIYTNKT
                     ETTIPYKIVKEINLINENLNDFPIYISLNTLQGVIYIVTKCSYVYVFDEGTLSLIVKE
                     KISEDNIFICCDSKNGEGIYAVNKKGIIFYITINYFHLINHLKLSNLEFKDKIIKNLC
                     VKYGYPGCDYISAYKKCINDMDFKKASKIICLMKNTKIFEEHSNNIAESILIMNRKNI
                     DIRIIPPLRTQQVLNSFKSFKNTSGQLSPLLLYFSVLLEYDKLNTYESIELVKPVVLQ
                     KKKEYLEKWIKDDKLTCSEELGDLVKVLDLRLALNIYLRCSSHNKILSTYCLLNMFNN
                     VLSYINNFKQINFDYVGIFITIVNYEFPQGNADNKDSAFVGDFNSGSKNETMDFFNDG
                     AASISNDQNNSINKLNKGHINEVAIQYIKFLCENNISFDINKIIDYLLSKKKLQEATS
                     ILLDYLKENKAEHKNLQTKLFEFNLYNNVQVAETLFQMDIFTYYDKNKIAYLCEEKGL
                     YQRALENYTNINDIKRVITKSTCFQKGGNNNNTTTTTTTTSNMGDGHFDMNNVSKGKI
                     SIEWIKNYFSTLSDSVCEELLFDFMKGSKINMEVVISICVQYYDKIGIKKIINKFEEN
                     KNYEGIFYFVSNILNDLPNLIVKGNNNDNLSNNSLSTYNNMNDEDASILLTSDICSTP
                     QYSYETANLKLEDVHYIMFKYIEACVKINNIQELDRICKDKNAKYNPEQIKNFLKDCK
                     LSDPRPLIYVCDIHNYIEELAEYLYKNSLLKYIEVYVIKVNPNNAHKVIGVLLDLDAS
                     EDFLLNLLNNIKNISNIGNLIEIAEKRNRLKLLLPWLESRANEGYENIELHNALAKIY
                     IDLNKDPENFLKNNNFYDKKMIGKYCEDLDPHLAYTAYERSNGECDEELIYITSKNGL
                     FKLQARYLISRQSIELWKVVLDETNKYRKNVIDQVIGSTLIESNNADEITVTVKAFIE
                     KKLSSELIELLEKIVLHNSEFSDNKNLQNLLILTAIKSDSKKVMEYINRLDNFSGPQI
                     AAVAYEYKLREEAFVIYKKFNCYTSAISVLLDKILYNKNQKNKNFHEFSSHDSTVPYY
                     SNNNNKDNIFGNKSDNENDEYLKQFTQNNSYREKYGSYYNNNDISNSNNNNIHLHPEF
                     NNEDKFSSDTLHNYEDIDSTESTCDLSNYEDDLNRAIEFAQKCNNNDVWFILGKAQLK
                     LNKIIDSIDSFVKSNNPEAYKEVIEKCKENNFYEHLITYLNTLREQNSLKDVLVDSEL
                     LYAYAKLKKTNEMNKFINTTNSANLQLIGDRLYKEEEYEVAKILYSNIPNNQKLTACY
                     LKLKEYALAIEAAKKAKSLKTWKEVNFICVKYKQLKYAHTAGLQLIMHADHLDEIIKI
                     YEKKKYINELMNLLENGLNNERAHVGIYTELGILYAKYKPEKLMEFIRNYTNKMNTRK
                     LIDVCENEYLLKEAVYLYISYDEYNLAVDTIIKHSPTAYTADTFMQVIHKVTNSDIIH
                     KVIDFYIEEHPLNLYNLLKILENKIDNNRLVQTLKKSNNLPLIQKYLEDIQAQNITAV
                     NETLNEIYLQNDDYISLRNSIDEYDNFNQTNLLNKLENHKLPEMRRIAALLYKKNKKY
                     KEAINLSKKEKQYKDAIEIARVSKNNVYIEDLINYFIESKNKEAFCACLIVCYDLLKP
                     DYILEIVWTSGFKDQAMLYFIQIISDYTNQIEHMKKQIEDMEKEKKMNKSAPNDYSAM
                     NNQFNYSLNNNLSIMPPQNNFMSSNSFDKYDMFNNNTHF"
     repeat_region   3056. .3105
                     /rpt_type=tandem
     repeat_region   4923. .4948
                     /rpt_type=tandem
     repeat_region   8613. .8642
                     /rpt_type=tandem
     repeat_region   8716. .8742
                     /rpt_type=tandem
     repeat_region   9080. .9130
                     /rpt_type=tandem
     gene            complement(<11613. .>19156)
                     /locus_tag="PFL0935c"
     mRNA            complement(join(<11613. .12941,13715. .>19156))
                     /locus_tag="PFL0935c"
     CDS             complement(join(11613. .12941,13715. .19156))
                     /locus_tag="PFL0935c"
                     /note="HMMPfam hit to PF03011, Plasmodium falciparum
                     erythrocyte membrane protein (PFEMP), score 1.3e-56"
```

```
                     /codon_start=1
                     /product="erythrocyte membrane protein 1 (PfEMP1)"
                     /protein_id="AAN36275.1"
                     /db_xref="GI:23496720"
                     /translation="MVAAAKGGGSSQDAKHVLDEIGKDVYETVEKDAKIYKEALKGNL
                     QHAKGMGERASSNKTCTLVKEYYEHFNGDANSNRYPCKELSGKMGENRFSDTLGGQCT
                     NKKIEGNKNNCGACAPYRRLHLCHHNLESIDTTSTTSDTLLAEVCYAAKFEGETLTTQ
                     HGQHQQTNPGTASQLCTVLARSFADIGDIVRGKDLFLGNDEEKKKRDELEENLQKIFA
                     KIHSGLTSTRGKQNGELQKRYKGDAKKNFYQLREDWWDANRHTVWEAITCGAGESDKY
                     FRQTACGTGTPTHKQCRCDDKANVDPPTYFDYVPQYLRWFEEWAEDFCRKKNKKLKDV
                     KTNCRGDSGTDRYCSGNGYDCTQTIRAIGKLVIGEHCTNCSVWCRLYEKWIDNQKLEF
                     LKQRNKYQTEISGGGASGGASGRSRKTRAATIKYEGYEKKFYEEFKKKGNYGTVDAFL
                     GLLNNEKACKDINEEKEKIDFTKNVEDDKNINKEGTFYHSQYCQPCPGCGMKKTKNGN
                     GWEEKDTNQCNIKLYKPTSSAKPTEIKILKSGEKHDDIKEKIEQFCKTQNGSDGSGGG
                     NSEKKELYDEWKCYQIGELTKDGQEGMDDPDYDRLVETGGGLCILENQKKKEKSDPKS
                     QNEPEQFQKTFNEFFYFWIGRFLNDSMYWRGKVNSCINKSKREKCEEECKKECECFQK
                     WIGKKKEEWDKIKEHFKTQDFGNDALLGPEMKSPSYVLKFVLNIDELFQNIKSGYGNV
                     KELKGINKILDEEKKREEEERADGVVGVASGGDNQNNTTIDKLLDHEDKDATKCKNCQ
                     EPTKPASKPEDLARSGASPDTPREDSPPADLDDEEHDEDDEDEDDDAEEEEEEKKEE
                     EEEHKEEENEAAPEEKKEEGSSSPEEVEEKAKEAPAGPDACNIVQTLFSSIDKFSDAC
                     TLKYGKTAPTGWKCVTPSGDKAATSSEGKGSSDGGAKDGVGVNGGALQRNKRDLATPS
                     AKSGDTTGGKDGATGKSDGSICVPPRRRRLYIQKLHDWAEKVGDTATQPQVDTPSQSD
                     KLRDAFIESAAVETFFLWHRYKKIKDKEKLEEQQRQRENGELPGLSSSGDGDSNDPQS
                     KLEKGEIPDGFLRQMFYTLADYKDILYSGSNDNTKSSTYNDIINGDKEIQEREKDIKG
                     AISTYFSNSGTTPTPPVKPSVTTPQTWWELHGPHIWNGMICALTYKENGIDKPEVDTA
                     VRAQLWDSGKNTPQNEKYQYTNVKLEEEASGAKINIPDTSGDNTPTLNTPTLKNFVEI
                     PTFFRWLHEWGSDFCRQRKRMLKNVKHNCRNIERGGHEYCGGDGHDCTRDGIQHNNML
                     ADPDCRDCHIQCRKYRKWIEKKFEEFHNQKNTYGKEKQKLNGNSNDGDNTKFFKEIEN
                     RSFDQFLTSLKHCKNNEGDGSDPKNKIDFKEPLKTFGPLEYCKTCPPNKVNCNGPSRR
                     SGGNDQCTAVNGNEWEKIFSENGGNSTTIDVHMIDRRGPYMEKKSQKLENSENPLFKT
                     SRLFKGIREQKWTCKFKDKNTDVCKLDKFDQEVDLNQYTTFKVLLIYWLEDFIEGYYL
                     LKKRKIIEKCTQKGEKTCDDESKNDCACVKKWVEKKTTEWEQIKEHFKNRNQKDGDGN
                     DMKSSVRQLLDPLIYRMDLANGKGKINELKEFLKSYECKCVDNAGNSEKDVVECLLQK
                     LETKAKKCKDPSSGENQKTPCQEYTPPDDDEDLLLEEENTVKAPTFCPKPPKKEEQTE
                     ETCDKAADEKVEETVAQNEENNIPPAGPAPPADSPPAGPATDSGKENVVPEPPAPAAP
                     PSPPIPPLATSTLAWSVGIGFAAFTYFYLKKKTKASVGNLFQILQIPKSDYDIPTLKS
                     KNRYIPYRSGTYKGKTYIYMEGDADEDKYTFMSDTTDITSSESEYEELDINDIYVPGS
                     PKYKTLIEVVLEPSGNNTTASGKNTPSDTQNDIQNDGIPSSKITDNEWNTLKDEFISN
                     MLQNQPKDVPNDYSSGDIPFNTQPNTLYFDNNQEKPFITSIHDRNLYTGEEYSYNVNM
                     VNNDNIPINRDNNPYSGIDLINDSLNSNKVDIYDELLKRKENELFGTEHHPKHTNIYN
                     VAKPARDDPITNQINLFHKWLDRHRDMCEKWDTNNKVDILNQLKEEWENDNSNSGNKT
                     SGNITPTSDIPSGKLSDIPSTNKMLNSDVSIQIHIDKPNQVDDNIYLDTYPDKYTVDN
                     INPVDTHTNPNLVGNINPVDQNSNLTFPSNPNPAYDNIYIDHNNEDLPSKVQIEMSVK
                     NGEMAKEK"
     repeat_region   12602. .12648
                     /rpt_type=tandem
     repeat_region   12960. .13070
                     /rpt_type=tandem
     repeat_region   12966. .13039
                     /rpt_type=tandem
     repeat_region   13048. .13072
                     /rpt_type=tandem
     repeat_region   13469. .13493
                     /rpt_type=tandem
     repeat_region   13562. .13586
                     /rpt_type=tandem
     repeat_region   16563. .16614
                     /rpt_type=tandem
     repeat_region   17915. .17943
                     /rpt_type=tandem
     repeat_region   19242. .19283
                     /rpt_type=tandem
     repeat_region   19334. .19422
                     /rpt_type=tandem
     repeat_region   20141. .20273
                     /rpt_type=tandem
     repeat_region   20940. .21079
                     /rpt_type=tandem
     repeat_region   21067. .21310

Alignment Scores:
Pred. No.:          48.3          Length:      252650
Score:              76.00         Matches:     11
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            90.48%        Indels:        0
DB:                     2             Gaps:          0

US-09-858-332G-15 (1-12) x AE014847 (1-252650)

Qy          2 AsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||
Db     169165 AATCATAATCATAATCATAATCATAATCATAAT 169197

RESULT 240
AC098369/c
LOCUS       AC098369               336001 bp    DNA     linear   HTG 13-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-2C20, *** SEQUENCING IN PROGRESS ***,
            25 unordered pieces.
ACCESSION   AC098369
VERSION     AC098369.5  GI:24941445
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 336001)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 336001)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
```

```
REFERENCE   3  (bases 1 to 336001)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 13, 2002 this sequence version replaced gi:23322421.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: TUNZ
                Center clone name: CH230-2C20
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 210771 bases at least Q40
                Consensus quality: 220796 bases at least Q30
                Consensus quality: 226806 bases at least Q20
                Estimated insert size: 204716; sum-of-contigs estimation
                Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 25 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     8804: contig of 8804 bp in length
            *    8805     8904: gap of unknown length
            *    8905   104226: contig of 95322 bp in length
            *  104227   104326: gap of unknown length
            *  104327   127899: contig of 23573 bp in length
            *  127900   127999: gap of unknown length
            *  128000   140167: contig of 12168 bp in length
            *  140168   140267: gap of unknown length
            *  140268   162385: contig of 22118 bp in length
            *  162386   162485: gap of unknown length
            *  162486   174547: contig of 12062 bp in length
            *  174548   174647: gap of unknown length
            *  174648   185645: contig of 10998 bp in length
            *  185646   185745: gap of unknown length
            *  185746   250087: contig of 64342 bp in length
            *  250088   250187: gap of unknown length
            *  250188   262001: contig of 11814 bp in length
            *  262002   262101: gap of unknown length
            *  262102   283121: contig of 21020 bp in length
            *  283122   283221: gap of unknown length
            *  283222   290841: contig of 7620 bp in length
            *  290842   290941: gap of unknown length
            *  290942   304995: contig of 14054 bp in length
            *  304996   305095: gap of unknown length
            *  305096   306168: contig of 1073 bp in length
            *  306169   306268: gap of unknown length
            *  306269   307329: contig of 1061 bp in length
            *  307330   307429: gap of unknown length
            *  307430   308940: contig of 1511 bp in length
            *  308941   309040: gap of unknown length
            *  309041   310157: contig of 1117 bp in length
            *  310158   310257: gap of unknown length
            *  310258   312333: contig of 2076 bp in length
            *  312334   312433: gap of unknown length
            *  312434   314495: contig of 2062 bp in length
            *  314496   314595: gap of unknown length
            *  314596   316232: contig of 1637 bp in length
            *  316233   316332: gap of unknown length
            *  316333   317361: contig of 1029 bp in length
            *  317362   317461: gap of unknown length
            *  317462   319107: contig of 1646 bp in length
            *  319108   319207: gap of unknown length
            *  319208   323347: contig of 4140 bp in length
            *  323348   323447: gap of unknown length
            *  323448   329675: contig of 6228 bp in length
            *  329676   329775: gap of unknown length
            *  329776   332269: contig of 2494 bp in length
            *  332270   332369: gap of unknown length
            *  332370   336001: contig of 3632 bp in length.
FEATURES             Location/Qualifiers
     source          1. .336001
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-2C20"
     misc_feature    1. .1761
                     /note="wgs_contig"
     gap             8805. .8904
                     /estimated_length=unknown
     misc_feature    32511. .33955
                     /note="wgs_contig"
     gap             104227. .104326
                     /estimated_length=unknown
     gap             127900. .127999
                     /estimated_length=unknown
     misc_feature    132568. .133554
                     /note="wgs_contig"
     misc_feature    138820. .140167
                     /note="wgs_contig"
     gap             140168. .140267
                     /estimated_length=unknown
     gap             162386. .162485
                     /estimated_length=unknown
     gap             174548. .174647
                     /estimated_length=unknown
     misc_feature    184453. .185645
                     /note="wgs_contig"
     gap             185646. .185745
                     /estimated_length=unknown
     misc_feature    185746. .187910
                     /note="wgs_contig"
     misc_feature    214189. .216136
                     /note="wgs_contig"
     misc_feature    220831. .222435
                     /note="wgs_contig"
     misc_feature    222937. .224602
                     /note="wgs_contig"
     misc_feature    248910. .250087
                     /note="wgs_contig"
     gap             250088. .250187
                     /estimated_length=unknown
     misc_feature    260990. .262001
                     /note="wgs_contig"
     gap             262002. .262101
                     /estimated_length=unknown

Alignment Scores:
Pred. No.:              63.6          Length:          336001
Score:                  76.00         Matches:         11
Percent Similarity:     100.00%       Conservative:    0
Best Local Similarity:  100.00%       Mismatches:      0
```

```
Query Match:            90.48%      Indels:        0
DB:                     14          Gaps:          0

US-09-858-332G-15 (1-12) x AC098369 (1-336001)

Qy          2 AsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||||||||
Db     254504 AATCACAATCATAATCACAATCATAATCATAAT 254472

RESULT 241
AC002367/c
LOCUS       AC002367               148600 bp    DNA     linear   PRI 28-MAR-2002
DEFINITION  Homo sapiens Xp22 PAC RPCI1-22N22 (Research Park PAC library)
            complete sequence.
ACCESSION   AC002367
VERSION     AC002367.1  GI:2811095
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 148600)
  AUTHORS   Muzny,D., Arenson,A.D., Brundage,E., Carvelli,K., Chen,E., Di,W.,
            Ding,Y., Dugan,S., Durbin,J., Forcum,J., Ganesh,R., Garcia,C.,
            Goodman,M., Gorrell,J.H., Haywood,M., Jackson,L., Kampal,R.,
            Karpathy,S., Leal,B., Liu,W., Logan,O., Lu,J., Ly,T., Martinez,C.,
            Oswal,G., Perez,L., Rashid,N.D., Rowland,K., Savage,L.,
            Scherer,S.S., Shen,H., Timms,K.M., Todd,J., Vo,Q., Worley,K.C.,
            Yu,W., Chinault,C., Nelson,D. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 148600)
  AUTHORS   Chiu,M.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1997) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 148600)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-1998) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 148600)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-2002) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jan 27, 1998 this sequence version replaced gi:2689462.
            Sequencing is completed to a minimum standard of double strand
            coverage with a minimum of 2 clones and 2 reads with no ambiguities
            or 2 chemistries with a minimum of 2 clones and 3 reads with no
            ambiguities. If the sequence quality does not meet this standard,
            it will be indicated in the annotation.

            The repeat regions shown were identified using RepeatMasker by
            Adrian Smit.

            Sequence similarities were identified using Powerblast by Jinghui
            Zhang.

            Exon/Intron boundaries of identified genes were chosen if there
            were canonical splice junctions that maintained sequence continuity
            across the splice junctions.
FEATURES             Location/Qualifiers
     source          1. .148600
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /clone="RPCI1-22N22"
     repeat_region   complement(30. .79)
                     /rpt_family="MER5A"
     repeat_region   complement(121. .297)
                     /rpt_family="MER20"
     repeat_region   611. .1081
                     /rpt_family="MLT1D"
     repeat_region   complement(1224. .1340)
                     /rpt_family="MIR"
     repeat_region   complement(1938. .2460)
                     /rpt_family="LINE2"
     repeat_region   2462. .2911
                     /rpt_family="MLT1C"
     repeat_region   complement(2925. .3293)
                     /rpt_family="LINE2"
     repeat_region   3308. .3795
                     /rpt_family="MLT1F"
     repeat_region   complement(3824. .4687)
                     /rpt_family="LINE2"
     repeat_region   4694. .4973
                     /rpt_family="AluJo"
     repeat_region   4977. .5003
                     /rpt_family="(GA)n"
     repeat_region   complement(5030. .5257)
                     /rpt_family="LINE2"
     repeat_region   complement(6108. .6259)
                     /rpt_family="AluYb8"
     repeat_region   complement(6320. .6369)
                     /rpt_family="AT_rich"
     repeat_region   7684. .7840
                     /rpt_family="MER69A"
     repeat_region   8880. .8935
                     /rpt_family="(CA)n"
     repeat_region   9214. .9309
                     /rpt_family="(TA)n"
     repeat_region   9309. .9388
                     /rpt_family="(GA)n"
     repeat_region   complement(10305. .10473)
                     /rpt_family="(CA)n"
     repeat_region   10794. .11667
                     /rpt_family="L1MA1"
     repeat_region   complement(11668. .11688)
                     /rpt_family="AT_rich"
     repeat_region   12572. .12659
                     /rpt_family="AT_rich"
     repeat_region   complement(12682. .12776)
                     /rpt_family="MER58B"
     repeat_region   complement(12777. .13220)
                     /rpt_family="L1PA5"
     repeat_region   complement(13234. .13268)
                     /rpt_family="(CA)n"
     repeat_region   complement(13524. .13643)
                     /rpt_family="AluSq/x"
     repeat_region   complement(13873. .13903)
                     /rpt_family="AT_rich"
     repeat_region   complement(15213. .15241)
                     /rpt_family="(GA)n"
     repeat_region   complement(15241. .15293)
                     /rpt_family="(CA)n"
     repeat_region   complement(15475. .15527)
                     /rpt_family="(CATA)n"
     repeat_region   complement(15985. .16190)
                     /rpt_family="MIR"
     repeat_region   complement(16322. .16954)
                     /rpt_family="L1PA16"
     repeat_region   complement(16956. .17259)
                     /rpt_family="AluSg"
     repeat_region   complement(17264. .17397)
                     /rpt_family="L1PA16"
     repeat_region   17886. .17924
                     /rpt_family="AT_rich"
     repeat_region   18008. .18045
                     /rpt_family="AT_rich"
     repeat_region   complement(18809. .19110)
                     /rpt_family="AluSq"
     repeat_region   complement(21003. .21140)
                     /rpt_family="MIR"
```

```
     repeat_region   21970. .22010
                     /rpt_family="(TAAA)n"
     repeat_region   22143. .22269
                     /rpt_family="MER5A"
     repeat_region   22532. .22723
                     /rpt_family="MER3"
     repeat_region   22770. .23074
                     /rpt_family="AluSq"
     repeat_region   23340. .23388
                     /rpt_family="(CATA)n"
     repeat_region   complement(24104. .24185)
                     /rpt_family="(GA)n"
     STS             24538. .24648
                     /standard_name="DXS1147"
     repeat_region   24821. .25039
                     /rpt_family="MER58A"
     repeat_region   complement(25936. .25966)
                     /rpt_family="AT_rich"
     repeat_region   complement(26086. .26214)
                     /rpt_family="MIR"
     repeat_region   complement(26381. .26421)
                     /rpt_family="LINE2"
     STS             26950. .27157
                     /standard_name="WI-5152, Chr. X, Homo sapiens G03871"
                     /db_xref="dbSTS:9357"
     repeat_region   complement(27506. .27536)
                     /rpt_family="AT_rich"
     repeat_region   27837. .28107
                     /rpt_family="AluSx"
     repeat_region   complement(28124. .28159)
                     /rpt_family="POLY_A"
     repeat_region   28160. .29415
                     /rpt_family="L1MA6"
     repeat_region   29499. .29525
                     /rpt_family="AT_rich"
     repeat_region   complement(30410. .30711)
                     /rpt_family="AluSx"
     repeat_region   complement(31829. .31979)
                     /rpt_family="L1PA4"
     repeat_region   31978. .32009
                     /rpt_family="L1PA5"
     repeat_region   complement(32087. .32117)
                     /rpt_family="AT_rich"
     repeat_region   32338. .32708
                     /rpt_family="LINE2"
     repeat_region   33194. .33426
                     /rpt_family="LINE2"
     repeat_region   33571. .33693
                     /rpt_family="MIR"
     repeat_region   33939. .34026
                     /rpt_family="(GAAA)n"
     repeat_region   34203. .34316
                     /rpt_family="L1MA7"
     repeat_region   36552. .37557
                     /rpt_family="L1PA15"
     repeat_region   complement(38009. .38156)
                     /rpt_family="LINE2"
     repeat_region   38434. .38651
                     /rpt_family="MIR"
     repeat_region   complement(38655. .38973)
                     /rpt_family="LINE2"
     repeat_region   39335. .39754
                     /rpt_family="LINE2"
     repeat_region   complement(40033. .40105)
                     /rpt_family="(GGAA)n"

Alignment Scores:
Pred. No.:              39              Length:         148600
Score:                  75.00           Matches:        11
Percent Similarity:     91.67%          Conservative:   0
Best Local Similarity:  91.67%          Mismatches:     1
Query Match:            89.29%          Indels:         0
DB:                     8               Gaps:           0
```

```
US-09-858-332G-15 (1-12) x AC002367 (1-148600)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||   ||||||||||||||||||||
Db      87281 CACAACCATAATTTTAACCATAATCATAATCATAAC 87246

RESULT 242
AC019165/c
LOCUS       AC019165              165738 bp    DNA     linear   HTG 07-JUL-2000
DEFINITION  Homo sapiens chromosome X clone RP11-149B9, WORKING DRAFT SEQUENCE,
            12 unordered pieces.
ACCESSION   AC019165
VERSION     AC019165.4  GI:8569824
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 165738)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 165738)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jun 16, 2000 this sequence version replaced gi:7230906.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0149B09
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 96%
            Sequencing vector: plasmid; 4%
            Chemistry: Dye-primer ET; 77% of reads
            Chemistry: Dye-terminator Big Dye; 23% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 159112 bases at least Q40
            Consensus quality: 160901 bases at least Q30
            Consensus quality: 162084 bases at least Q20
            Insert size: 188000; agarose-fp
            Insert size: 164638; sum-of-contigs
            Quality coverage: 4.32 in Q20 bases; agarose-fp
            Quality coverage: 4.90 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1412: contig of 1412 bp in length
            *     1413     1512: gap of unknown length
            *     1513     3280: contig of 1768 bp in length
            *     3281     3380: gap of unknown length
            *     3381     6343: contig of 2963 bp in length
            *     6344     6443: gap of unknown length
            *     6444    13485: contig of 7042 bp in length
            *    13486    13585: gap of unknown length
            *    13586    22798: contig of 9213 bp in length
            *    22799    22898: gap of unknown length
            *    22899    33955: contig of 11057 bp in length
            *    33956    34055: gap of unknown length
            *    34056    49624: contig of 15569 bp in length
```

```
            *    49625    49724: gap of unknown length
            *    49725    64591: contig of 14867 bp in length
            *    64592    64691: gap of unknown length
            *    64692    79609: contig of 14918 bp in length
            *    79610    79709: gap of unknown length
            *    79710   105126: contig of 25417 bp in length
            *   105127   105226: gap of unknown length
            *   105227   125931: contig of 20705 bp in length
            *   125932   126031: gap of unknown length
            *   126032   165738: contig of 39707 bp in length.
FEATURES             Location/Qualifiers
     source          1. .165738
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /clone="RP11-149B9"
     misc_feature    1. .1412
                     /note="assembly_name:Contig9"
     gap             1413. .1512
                     /estimated_length=unknown
     misc_feature    1513. .3280
                     /note="assembly_name:Contig10"
     gap             3281. .3380
                     /estimated_length=unknown
     misc_feature    3381. .6343
                     /note="assembly_name:Contig11"
     gap             6344. .6443
                     /estimated_length=unknown
     misc_feature    6444. .13485
                     /note="assembly_name:Contig12"
     gap             13486. .13585
                     /estimated_length=unknown
     misc_feature    13586. .22798
                     /note="assembly_name:Contig13"
     gap             22799. .22898
                     /estimated_length=unknown
     misc_feature    22899. .33955
                     /note="assembly_name:Contig14"
     gap             33956. .34055
                     /estimated_length=unknown
     misc_feature    34056. .49624
                     /note="assembly_name:Contig15"
     gap             49625. .49724
                     /estimated_length=unknown
     misc_feature    49725. .64591
                     /note="assembly_name:Contig16
                     clone_end:SP6
                     vector_side:left"
     gap             64592. .64691
                     /estimated_length=unknown
     misc_feature    64692. .79609
                     /note="assembly_name:Contig17"
     gap             79610. .79709
                     /estimated_length=unknown
     misc_feature    79710. .105126
                     /note="assembly_name:Contig18"
     gap             105127. .105226
                     /estimated_length=unknown
     misc_feature    105227. .125931
                     /note="assembly_name:Contig19"
     gap             125932. .126031
                     /estimated_length=unknown
     misc_feature    126032. .165738
                     /note="assembly_name:Contig20"
ORIGIN

Alignment Scores:
Pred. No.:              43.4          Length:         165738
Score:                  75.00         Matches:        11
Percent Similarity:     91.67%        Conservative:   0
Best Local Similarity:  91.67%        Mismatches:     1
Query Match:            89.29%        Indels:         0

DB:                     14            Gaps:           0

US-09-858-332G-15 (1-12) x AC019165 (1-165738)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||  ||||||||||||||||||||
Db      64428 CACAACCATAATTTTAACCATAATCATAATCATAAC 64393

RESULT 243
AC150208
LOCUS       AC150208              194447 bp    DNA     linear   PRI 08-JUL-2004
DEFINITION  Pan troglodytes chromosome X clone PTB-043C05 map human ortholog
            p21.3, complete sequence.
ACCESSION   AC150208
VERSION     AC150208.1  GI:50057851
KEYWORDS    HTG.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1  (bases 1 to 194447)
  AUTHORS   Taudien,S., Taenzer,S., Baumgart,C., Mueller,O. and Platzer,M.
  TITLE     DNA sequence of chimpanzee chromosome X
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 194447)
  AUTHORS   Lagemann,D. and Platzer,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUL-2004) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstr. 11, Jena, Thuringia 07745, Germany
COMMENT     ---------- Genome Center
                Center: Institute of Molecular Biotechnology
                Center code: IMB
                Web site: http://genome.imb-jena.de/
                Contact: gscj-submit@genome.imb-jena.de
            ---------- Project Information
                Center project name: cX34
                Center clone name: PTB-043C05
            ---------- Summary Statistics
                Sequencing clone based: pUC18; 82% of reads
                Whole genome shotgun: 18% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 194442 bases at least Q40
                Consensus quality: 194447 bases at least Q30
                Consensus quality: 194447 bases at least Q20
                Quality coverage: 12.99x
            ----------
            This sequence was finished with the support of external Whole
            genome shotgun data (http://www.ncbi.nlm.nih.gov/Traces/trace.fcgi)
            as follows unless otherwise noted: all regions were double
            stranded, sequenced with an alternate chemistry, or covered by high
            quality data (i.e., phred quality >= 30); an attempt was made to
            resolve all sequencing problems, such as compressions and repeats;
            all regions were covered by at least one plasmid subclone or more
            than one M13 subclone; and the assembly was confirmed by
            restriction digest.
            ----------
            Neighboring sequence information:
            This clone is overlapped by PTB-085O24.
            ----------
            Source information:
            The PTB1 chimpanzee BAC library was prepared from DNA isolated
            from cultured cells established from the blood of a single male
            chimpanzee. The clone may be obtained from Asao Fujiyama and
            co-workers (http://www.gsc.riken.go.jp).
            VECTOR:  pKS145
            ----------
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
```

```
            1 error in 10,000 bp.
            Base-by-base quality values are not generally visible from the
            GenBank flat file format but are available as part
            of this entry's ASN.1 file.
           ----------------------.
FEATURES             Location/Qualifiers
     source          1. .194447
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /chromosome="X"
                     /clone="PTB-043C05"
     source          58540. .194447
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-085O24"
                     /note="overlapping clone"
     misc_feature    3562. .3618
                     /note="single stranded/single chemistry region"
     misc_feature    11163. .11332
                     /note="single stranded/single chemistry region"
     misc_feature    11559. .11645
                     /note="single stranded/single chemistry region"
     misc_feature    34730. .35176
                     /note="single stranded/single chemistry region"
     misc_feature    35177. .35219
                     /note="single clone coverage , whole-genome shotgun only"
     misc_feature    37212. .37676
                     /note="single stranded/single chemistry region"
     misc_feature    38465. .38780
                     /note="single stranded/single chemistry region"
     misc_feature    55892. .56060
                     /note="single stranded/single chemistry region"
     misc_feature    85545. .85745
                     /note="single stranded/single chemistry region"
     misc_feature    92987. .93251
                     /note="single stranded/single chemistry region"
     misc_feature    110184. .110781
                     /note="single stranded/single chemistry region"
     misc_feature    134438. .134730
                     /note="single stranded/single chemistry region"
     misc_feature    142914. .143387
                     /note="single stranded/single chemistry region"
     misc_feature    143497. .143581
                     /note="whole-genome shotgun only"
     misc_feature    162529. .162742
                     /note="single stranded/single chemistry region"
     misc_feature    163420. .163909
                     /note="single stranded/single chemistry region"
ORIGIN

Alignment Scores:
Pred. No.:              50.6          Length:        194447
Score:                  75.00         Matches:       11
Percent Similarity:     91.67%        Conservative:  0
Best Local Similarity:  91.67%        Mismatches:    1
Query Match:            89.29%        Indels:        0
DB:                     8             Gaps:          0

US-09-858-332G-15 (1-12) x AC150208 (1-194447)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||   ||||||||||||||||||||
Db     104688 CACAACCATAATTTTAACCATAATCATAATCATAAC 104723

RESULT 244
AC150209
LOCUS       AC150209              207665 bp    DNA     linear   PRI 08-JUL-2004
DEFINITION  Pan troglodytes chromosome X clone PTB-085O24 map human ortholog
            p21.3, complete sequence.
ACCESSION   AC150209
VERSION     AC150209.1  GI:50057852
KEYWORDS    HTG.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1  (bases 1 to 207665)
  AUTHORS   Taudien,S., Taenzer,S., Baumgart,C., Mueller,O. and Platzer,M.
  TITLE     DNA sequence of chimpanzee chromosome X
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 207665)
  AUTHORS   Lagemann,D. and Platzer,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUL-2004) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstr. 11, Jena, Thuringia 07745, Germany
COMMENT     -------------- Genome Center
                 Center: Institute of Molecular Biotechnology
                 Center code: IMB
                 Web site: http://genome.imb-jena.de/
                 Contact: gscj-submit@genome.imb-jena.de
            -------------- Project Information
                 Center project name: cX42
                 Center clone name: PTB-085O24
            -------------- Summary Statistics
                 Sequencing clone based: pUC18; 82% of reads
                 Whole genome shotgun: 18% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990329
                 Consensus quality: 207648 bases at least Q40
                 Consensus quality: 207664 bases at least Q30
                 Consensus quality: 207665 bases at least Q20
                 Quality coverage: 12.17x
            --------------
            This sequence was finished with the support of external Whole
            genome shotgun data (http://www.ncbi.nlm.nih.gov/Traces/trace.fcgi)
            as follows unless otherwise noted: all regions were double
            stranded, sequenced with an alternate chemistry, or covered by high
            quality data (i.e., phred quality >= 30); an attempt was made to
            resolve all sequencing problems, such as compressions and repeats;
            all regions were covered by at least one plasmid subclone or more
            than one M13 subclone; and the assembly was confirmed by
            restriction digest.
            --------------
            Neighboring sequence information:
            This clone is overlapped by PTB-043C05.
            --------------
            Source information:
            The PTB1 chimpanzee BAC library was prepared from DNA isolated
            from cultured cells established from the blood of a single male
            chimpanzee. The clone may be obtained from Asao Fujiyama and
            co-workers (http://www.gsc.riken.go.jp).
            VECTOR:  pKS145
            --------------
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.
            ----------------------.
FEATURES             Location/Qualifiers
     source          1. .207665
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /chromosome="X"
                     /clone="PTB-085O24"
     source          1. .135906
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
```

```
                     /db_xref="taxon:9598"
                     /clone="PTB-043C05"
                     /note="overlapping clone"
     misc_feature    27006. .27206
                     /note="single stranded/single chemistry region"
     misc_feature    34448. .34712
                     /note="single stranded/single chemistry region"
     misc_feature    51645. .52242
                     /note="single stranded/single chemistry region"
     misc_feature    75899. .76191
                     /note="single stranded/single chemistry region"
     misc_feature    84375. .84848
                     /note="single stranded/single chemistry region"
     misc_feature    84958. .85042
                     /note="whole-genome shotgun only"
     misc_feature    103990. .104203
                     /note="single stranded/single chemistry region"
     misc_feature    104881. .105370
                     /note="single stranded/single chemistry region"
     misc_feature    138253. .138545
                     /note="single stranded/single chemistry region"
     misc_feature    139383. .139638
                     /note="single stranded/single chemistry region"
     misc_feature    147881. .148396
                     /note="single stranded/single chemistry region"
     misc_feature    165833. .165860
                     /note="single stranded/single chemistry region"
     misc_feature    170374. .170503
                     /note="single stranded/single chemistry region"
     misc_feature    171832. .171886
                     /note="single stranded/single chemistry region"
     misc_feature    176755. .176935
                     /note="single stranded/single chemistry region"
     misc_feature    184529. .184862
                     /note="single stranded/single chemistry region"
     misc_feature    185925. .186387
                     /note="single stranded/single chemistry region"
     misc_feature    188504. .188773
                     /note="single stranded/single chemistry region"
     misc_feature    191175. .191212
                     /note="single clone coverage"
     misc_feature    197391. .197429
                     /note="single stranded/single chemistry region"
     misc_feature    198026. .198174
                     /note="single stranded/single chemistry region"
     misc_feature    198026. .198042
                     /note="whole-genome shotgun only"
     misc_feature    199330. .199430
                     /note="single stranded/single chemistry region"
     misc_feature    200150. .200252
                     /note="single stranded/single chemistry region"
     misc_feature    200150
                     /note="low quality region"
     misc_feature    200265. .200530
                     /note="pcr product sequence only"
ORIGIN

Alignment Scores:
Pred. No.:                53.9          Length:         207665
Score:                    75.00         Matches:        11
Percent Similarity:       91.67%        Conservative:   0
Best Local Similarity:    91.67%        Mismatches:     1
Query Match:              89.29%        Indels:         0
DB:                       8             Gaps:           0

US-09-858-332G-15 (1-12) x AC150209 (1-207665)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||   ||||||||||||||||||||||
Db      46149 CACAACCATAATTTTAACCATAATCATAATCATAAC 46184

RESULT 245
DVNEURS1
LOCUS       DVNEURS1                 465 bp    DNA     linear   INV 23-AUG-2002
DEFINITION  Drosophila virilis neuralized protein (neu) gene, exon 1.
ACCESSION   U12591
VERSION     U12591.1  GI:535310
KEYWORDS    .
SEGMENT     1 of 3
SOURCE      Drosophila virilis
  ORGANISM  Drosophila virilis
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 465)
  AUTHORS   Zhou,L. and Boulianne,G.L.
  TITLE     Comparison of the neuralized genes of Drosophila virilis and D.
            melanogaster
  JOURNAL   Genome 37 (5), 840-847 (1994)
   PUBMED   8001814
REFERENCE   2  (bases 1 to 465)
  AUTHORS   Boulianne,G.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-1994) Gabrielle L. Boulianne, Physiology,
            University of Toronto, 6 Queens Park Crescent, Toronto, Ontario M5S
            1A8, Canada
FEATURES             Location/Qualifiers
     source          1. .465
                     /organism="Drosophila virilis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7244"
                     /clone="dv5"
                     /clone_lib="Drosophila virilis sau3a genomic library; S.J.
                     Neufeld et al., J. Mol. Evol. 32:415-420(1991)."
     exon            <1. .332
                     /gene="neu"
                     /number=1
ORIGIN

Alignment Scores:
Pred. No.:                0.204         Length:         465
Score:                    74.00         Matches:        10
Percent Similarity:       100.00%       Conservative:   2
Best Local Similarity:    83.33%        Mismatches:     0
Query Match:              88.10%        Indels:         0
DB:                       2             Gaps:           0

US-09-858-332G-15 (1-12) x DVNEURS1 (1-465)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::||||||||||||||||||||||||
Db        161 CATCATCATCATCACAATCATAATCATAATCATAAT 196

RESULT 246
BV004681
LOCUS       BV004681                 467 bp    DNA     linear   STS 07-SEP-2002
DEFINITION  S208P6706FA10.T0 129S1/SvImJ Mus musculus STS genomic, sequence
            tagged site.
ACCESSION   BV004681
VERSION     BV004681.1  GI:22757757
KEYWORDS    STS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 467)
  AUTHORS   Wade,C.
  TITLE     Polymorphism Structure in the Mouse
  JOURNAL   Unpublished (2002)
COMMENT
            Contact: Kerstin Lindblad-Toh
            Whitehead Institute for Biomedical Research, Center for Genome
            Research
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172521477
```

```
          Fax: 6172580903
          Email: kersli@genome.wi.mit.edu
          Primer A: None
          Primer B: None
          STS size: 467
          Protocol:
          WGS-discovery: Paired-end low-coverage whole genome shotgun reads
          were generated from 129S1/SvImJ, C3H/HeJ, and BALB/cByJ. The WGS
          reads were placed uniquely on the MGSCv3 C57BL/6J assembly and SNP
          detection was carried out by SSAHA-SNP. 225,000 reads were
          annotated
          as STSs and 81,000 SNPs were annotated with alleles from C57BL/6J
          and the strain from which the particular read came. The validation
          rate for these SNPs was estimated at approximately 98%.
FEATURES             Location/Qualifiers
     source          1. .467
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="129S1/SvImJ"
                     /db_xref="taxon:10090"
                     /map="+ 9 11-411 114482865-114483253"
                     /clone_lib="129S1/SvImJ"
     STS             <1. .>467
ORIGIN

Alignment Scores:
Pred. No.:                0.204          Length:         467
Score:                    74.00          Matches:        10
Percent Similarity:       100.00%        Conservative:   2
Best Local Similarity:    83.33%         Mismatches:     0
Query Match:              88.10%         Indels:         0
DB:                       10             Gaps:           0

US-09-858-332G-15 (1-12) x BV004681 (1-467)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::||||||||||||||||||||||||||:::
Db        114 CACCATCACAACCACAACCACAACCACAACCACCAC 149

RESULT 247
AR546771
LOCUS       AR546771                747 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 1902 from patent US 6747137.
ACCESSION   AR546771
VERSION     AR546771.1  GI:53939946
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 747)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 1902 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .747
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:                0.321          Length:         747
Score:                    74.00          Matches:        10
Percent Similarity:       100.00%        Conservative:   2
Best Local Similarity:    83.33%         Mismatches:     0
Query Match:              88.10%         Indels:         0
DB:                       6              Gaps:           0

US-09-858-332G-15 (1-12) x AR546771 (1-747)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||||||||||||||||||||||||||:::|||:::

Db         91 CATAATCACAACCACAATCACAATCATCATCACCAT 126

RESULT 248
AY122262
LOCUS       AY122262                5687 bp    mRNA    linear   INV 19-JUN-2002
DEFINITION  Drosophila melanogaster SD13096 full insert cDNA.
ACCESSION   AY122262
VERSION     AY122262.1  GI:21483597
KEYWORDS    FLI_CDNA.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 5687)
  AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J.,
            Champe,M., Chavez,C., Dorsett,V., Dresnek,D., Farfan,D., Frise,E.,
            George,R., Gonzalez,M., Guarin,H., Kronmiller,B., Li,P., Liao,G.,
            Miranda,A., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
            Patel,S., Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Rubin,G.M.
            and Celniker,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2002) Berkeley Drosophila Genome Project,
            Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, USA
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
             Lawrence Berkeley National Laboratory
            Berkeley, CA 94720
             This clone was sequenced as part of a high-throughput process to
            sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
            Science 2000). The sequence has been subjected to integrity checks
            for sequence accuracy, presence of a polyA tail and contiguity
            within 100 kb in the genome. Thus we believe the sequence to
            reflect accurately this particular cDNA clone. However, there are
            artifacts associated with the generation of cDNA clones that may
            have not been detected in our initial analyses such as internal
            priming, priming from contaminating genomic DNA, retained introns
            due to reverse transcription of unspliced precursor RNAs, and
            reverse transcriptase errors that result in single base changes.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our Web site
            (http://fruitfly.berkeley.edu) or send email to
            cdna@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1. .5687
                     /organism="Drosophila melanogaster"
                     /mol_type="mRNA"
                     /db_xref="taxon:7227"
     gene            1. .5687
                     /gene="CG10231"
                     /db_xref="FLYBASE:FBgn0032686"
     CDS             411. .5045
                     /gene="CG10231"
                     /note="Longest ORF"
                     /codon_start=1
                     /product="SD13096p"
                     /protein_id="AAM52774.1"
                     /db_xref="GI:21483598"
                     /db_xref="FLYBASE:FBgn0032686"
                     /translation="MKVTQSEENTRNTSDRSKSVQTNRKFDNFNWLLSCGLLAAVKST
                     KLIPQLPRQQQPKKYNLRYAAELLKFDKIQFIKTEGLTFALLAGPMGQAASMCRFRGC
                     RYKNKNKSSKQQQQQQQQQQQQQQQQHQQQQQQTPSHSPQIQHHSEIIPATTGLHLR
                     SIEEPATTPLQFQPTGRMNTEQGGTGYGGYGSSEHSLLIATRHAGVPLPLAQHQPLPA
                     HYQPLNHSGAAPPSSSNGSSSSGGGVQTSATPQQQQQQYQVQQPYQYQYQHHYHHQANS
                     PQHHRPYDPEHARMEAWLDENQEFVQDYFIRKATRQTVDAWLVSHATSAGNDVVSSTS
                     PTHANGQTSSSRGGSGATTPVRKISAHEFERGGLLKPIVNTIDGTPTFLSIGPPMDNG
                     SVGGSCSNLQNVGGVVAGQYQYNHQQHHNHAHLHHSQHSHYQAGGAVGSSSLGSTGG
                     ASGAGGAPSLGGSGGAGNGHQYPYYHCHQRPQRLSRNELKQLDEKELIFELVKDICNE
                     LEVRTLCHKILQNVSILLNADRGSLFLVQGRCNGPDGLKKCLVSKLFDVCPRSTVEEM
                     EQQDEVRVAWGTGIAGHVAESGEPVNIPDAYQDERFNCEIDSLTGYRTKALLCMPIKD
                     SSGDVIGVAQVINKMNGECFSEIDEKVFSSYLQFCGIGLRNAQLYEKSQLEIKRNQVL
                     LDLARMIFEEQSTIEHMVFRILTHMQSLIQCQRVQILLVHEADKGSFSRVFDFEANDL
```

SEEEATSRTSPYESRFPINIGITGHVATTGETVNVPNAYEDDRFDASVDENSCFKHRS
ILCMAIKNSLGQIIGVIQLINKFNELDFTKNDENFVEAFAIFCGMGIHNTHMYEKAIV
AMAKQSVTLEVLSYHASATMDEAHRLRRLRVPSAVHFRLHDFKFDDIHFEDDDTLKAC
LRMFLDLDFVERFHIDYEVLCRWLLSVKKNYRNVTYHNWRHAFNVAQMMFAILTTTQW
WKIFGEIECLALIIGCLCHDLDHRGTNNSFQIKASSPLAQLYSTSTMEHHHFDQCLMI
LNSPGNQILANLSSDDYCRVIRVLEDAILSTDLAVYFKKRGPFLESVSQPTSYWVAEE
PRALLRAMSMTVCDLSAITKPWEIEKRVADLVSSEFFEQGDMEKQELNITPIDIMNRE
KEDELPMMQVNFIDSICLPIYEAFATLSDKLEPLVEGVRDNRGHWIDLADVVKTKTSQ
DQEPEEEQQQQNVISNGDCKAMSDDDVAASEAEVAVDSPSEKASVNGSNVANNSSNTN
KKIAVASHPTSTQPSDDDNDADADVDDVDEQAAEENGHDAEVDEASCRSNSTCSSSTA
SSRLSTPPPTGEDDSTPVSPLKTLQAKLVAANLNALQRQTSNQAQTQKQRCKSCDHSR
SGLQVRKTSSLRGAQELDLDSKTRNGTHAALCKSTPVINNHSHHHNHSHSNHNHNHH
HHHHHHSHHNHSQHGIGIGSASIGGSGLISLTTPLLAMDSDRIPKIVGKIGNLDGLPF
ANGIGGPQNGHGLPFGSYQHHHHHQHHHHLLARRHSETNSNGATAMAVEK"

ORIGIN

Alignment Scores:

Pred. No.:	2.27	Length:	5687
Score:	74.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	83.33%	Mismatches:	0
Query Match:	88.10%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-15 (1-12) x AY122262 (1-5687)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||||||||:::|||:::||||||||||||||||||
Db       4677 CATAATCATAGTCACAGCCACAACCACAACCACAAC 4712
```

RESULT 249
AF474341/c
LOCUS AF474341 8167 bp DNA linear INV 06-MAR-2002
DEFINITION Dictyostelium discoideum ABC transporter ABCC.9 (abcC9) gene, complete cds.
ACCESSION AF474341
VERSION AF474341.1 GI:19172039
KEYWORDS .
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 8167)
AUTHORS Anjard,C. and Loomis,W.F.
TITLE Evolution of the ABC transporters of Dictyostelium
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 8167)
AUTHORS Anjard,C. and Loomis,W.F.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-2002) Biology Dept 0368, UCSD, 9500 Gilman drive, La Jolla, CA 92093, USA
FEATURES Location/Qualifiers
source 1. .8167
/organism="Dictyostelium discoideum"
/mol_type="genomic DNA"
/strain="Ax4"
/db_xref="taxon:44689"
gene <320. .>4553
/gene="abcC9"
mRNA join(<320. .792,989. .>4553)
/gene="abcC9"
/product="ABC transporter ABCC.9"
CDS join(320. .792,989. .4553)
/gene="abcC9"
/note="protein family C"
/codon_start=1
/product="ABC transporter ABCC.9"
/protein_id="AAL85712.1"
/db_xref="GI:19172040"
/translation="MIRNKIKNKKSKSDYYSEISLEEEEDKEEEIKDKKVLSKEEIKE
NENKNEKIKQTQCPEDNASLFSKITFGWITKLIVKAYIKKSLDIKDIFDVPNYLKSNY
TSKFLTTEKQTFKRNTKYSLILNIYINFILLKNKKLILIQFLRVLFTLLSPLILKLFI
EFTQRIDSEKSIIEGIIYCGLLFFSSFCSSLVDEYLVWFGMITSSQVKSCLTCLIPEK
SLKLSNSAKVKYNAAKITNLISIDVDNFSNFFWSNSVEIFFQPFQLIFLLILLIPEVG

WSGFVGTAVILISFPINSYFGKKTSDYYEKLLKYTDKRVSTTSEFINGIRFIKMYAWE
SLFLKKIEKQRKQELKILLERGLFWLGQIIIINTNSTFVFVATMVTYSLSGNKMKLET
AFTAMNILDSIRILLIVMPYCYYSIMELIPSNKRIEKFLSTLEIQQNLQTTKNSNTIS
INNGTFKWKDENFEETDDDDDHDDDDDEGEKEIEGKEEKEEINNFIFENINFKAPIGK
LTMICSPVGSGKTSLINALIGEIEKVNGEINGVPENISFTSQQPFLLSTSLRENILFG
KELDIEYYKQVLDACCLVSDLTQLSALDLTEIGEGGINLSGGQKQRVSLARALYSNSD
FILMDEPLSAVDPEVANHLFEKCIQGMMKNKTRILVTHQIQFIPYADHIVIIKDGKII
QGTYKELKDENGIDFESIIKTKNSNLNLNKTIKDEEEKEDKEGEEKEEDKKEMILKLI
KINYSDELKEKAKLLVEEDRNEGEVSFKTYKEYFYYGSSNLFLFITLLVFLIGLIVNR
LSDFWLTIWTEQSFKEKSQTFYIICFVSIFIVSLILIFIRYSLVAKIGFKSSKKLHDL
LINSISKSPIQFFDSNPSGRILNRFSNDISDIDTSMIDMFSDTLEYIFAIVVGIFSII
YINPMTIVPFLILSMFYYQIFNIYRVSVRELNRCKSISQSPIISFLSECCNGLSTIRA
FKQQSRFIGIMNDNIDKNLKCEFASFAVEMWISIRLELLSSIIVFIVSLFSLFNGYSN
SALSILSVSTARSITSYLKSACRQMVDLERKMNSFERVNNYIKLPSEGSGSGVKSMEF
ISEKDLLNWPEKGNIQFNNVQVNYNSNSVPSLKDITFSVESNEKIGIVGRTGAGKSTI
ANCLFRLVECSKGSILIDGIDIKTIDLNKLRGSLGIVPQEPWLFSGTIRSNIDPLNQY
DDEMIWNYLEMVKLKKLIIEMPLKLNSKIHENGNTTLSYGQKQLLCLCRCLIKNPKLI
IMDEATSSVDFQTAETIKSVINENLVNNTILTIAHRLDIIIDSDKIAVIDNSKLIEFE
NPKNLINSNSKFRKIVNFQTKIQTN"

ORIGIN

Alignment Scores:

Pred. No.:	3.22	Length:	8167
Score:	74.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	83.33%	Mismatches:	0
Query Match:	88.10%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-15 (1-12) x AF474341 (1-8167)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db       7419 CATAGTCATAGTCATAACCATAATCATAATCACAAT 7384
```

RESULT 250
NCB12N19
LOCUS NCB12N19 42273 bp DNA linear PLN 18-JAN-2002
DEFINITION Neurospora crassa DNA linkage group II BAC clone B12N19.
ACCESSION AL669987
VERSION AL669987.1 GI:18376315
KEYWORDS .
SOURCE Neurospora crassa
ORGANISM Neurospora crassa
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
REFERENCE 1
AUTHORS Schulte,U., Aign,V., Hoheisel,J., Brandt,P., Fartmann,B., Holland,R., Nyakatura,G., Mewes,H.W. and Mannhaupt,G.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 42273)
AUTHORS German Neurospora genome,project.
TITLE Direct Submission
JOURNAL Submitted (18-JAN-2002) MIPS, Institut fuer Bioinformatik, GSF-Forschungszentrum fuer Umwelt und Gesundheit, GmbH, Ingolstaedter Landstrasse 1, D-85764 Neuherberg, FRG, E-mail: G.Mannhaupt@gsf.de Project Coordinator: Ulrich Schulte, Institute of Biochemistry, Heinrich-Heine-University, D-40225 Duesseldorf, E-mail: ulrich-schulte@uni-duesseldorf.de
COMMENT BAC clone 12N19 (strain OR74A) is available at the Fungal Genetic Stock Center, http://www.fgsc.net
Sequencing was performed by MWG Biotech AG, Ebersberg, Germany, http://www.mwgdna.com
Information on performance of analysis and a more detailed annotation of this entry and other sequences of linkage groups II and V can be viewed at: http://mips.gsf.de/proj/neurospora.
FEATURES Location/Qualifiers
source 1. .42273
/organism="Neurospora crassa"
/mol_type="genomic DNA"
/db_xref="taxon:5141"
/chromosome="6"
misc_feature 1. .7212

```
                /note="overlap to BAC B11E6, EMBL:AL442043, please refer
                to this entry for analysis and annotation"
     gene       9830. .11707
                /gene="B12N19.020"
     CDS        join(9830. .10364,10432. .11357,11450. .11707)
                /gene="B12N19.020"
                /note="differs in three residues when compared to
                PIR:T47219
                Contains Amino acid permeases signature
                [GGPASLLIDFAIIGVMMFNVVYALGELAVMFP]
                contains EST gb:BF739557, BF739423"
                /codon_start=1
                /product="amino acid permease NAAP1"
                /protein_id="CAD21063.1"
                /db_xref="GI:18376316"
                /db_xref="GOA:Q8X0I5"
                /db_xref="UniProt/TrEMBL:Q8X0I5"
                /translation="MSDIEKGSAGSHDKETRVDAATLPAYEAGSVAEAQQEDFMTRNG
                LNLASFKPRRFHTGDGTEIERTMKSRHLHMIAIGGSIGAGFFVGSGKALHNGGPASLL
                IDFAIIGVMMFNVVYALGELAVMFPVTGGFYTYSTRFIDPSWGFAMGWNYVFQWAIVL
                PLELTVCGFTVGYWNQDITVWITIFLVAIIINVFGTLGYAEEEFWASAFKLGATIVF
                LIICLVMVLGGGPKGSRYDHYTGAALWYSPGAPANGFRGFCSCFVTAAFAFSGTELVG
                LAAAEAKNPVKSLPGAIKQVFWRITLFYILGLFFVGLLIPYDDENLLGANPFINVNAS
                PFVLVGKYANLTGFDSFMNVVILVSVLSLGVSCVYGGSRTLTALAQQGYAPKIFTYVD
                RSNRPLFSVLLIIAFGPLAYVNLAAAGPTVFDWLLSLSGLAALFTWGSICLSHIRFRK
                AWALKGHSVEEIPFRAAGGVWGSWLGLVLVVLVLIAQFYVAICPPGGGFNTVEGFFVQ
                YLALPVVIVFWIGGFIWKRTSWLRVDQIDVDTGRREHDWESINAYQAELAAMPAWKRW
                FHKLFV"
     exon       9830. .10364
                /gene="B12N19.020"
                /number=1
     intron     10365. .10431
                /gene="B12N19.020"
                /number=1
     exon       10432. .11357
                /gene="B12N19.020"
                /number=2
     intron     11358. .11449
                /gene="B12N19.020"
                /number=2
     repeat_region 11367. .11389
                /note="23 bp poly(t) tandem repeat"
     exon       11450. .11707
                /gene="B12N19.020"
                /number=3
     repeat_region 13995. .14030
                /note="36 bp atatatatatatatatatatatatat tandem repeat"
     gene       14175. .14909
                /gene="B12N19.030"
     CDS        14175. .14909
                /gene="B12N19.030"
                /note="contains EST gb:BG279737, BG279736"
                /codon_start=1
                /product="putative protein"
                /protein_id="CAD21064.1"
                /db_xref="GI:18376317"
                /db_xref="UniProt/TrEMBL:Q8X0I4"
                /translation="MGIFDYLRPRRQNGIPRKHTMPPPDILSYQNDALDKRHSKLSSM
                SPNAEQSQANTRPTTSSSWTSAITSTTTSTSISTSTSTSILTPTTTSHTTTTTPLSPP
                HSNYTLSNHDLNLNRPPTPPSRHPPPPSPPPSPPLFPTSEFTLHLPPPPPPPYARLSS
                RRTDLPLPSYAELVQERQIVPPRPRPLPTPREPPPSFEAAEIQEHDELHGMDPSSWPG
                CEYCRVYELWVSYWRGEGVGGVGVSG"
     exon       14175. .14909
                /gene="B12N19.030"
                /number=1
     gene       15344. .16676
                /gene="B12N19.040"
     gene       complement(join(15344. .16586,16639. .16676))
                /gene="B12N19.040"
     CDS        complement(join(15344. .16586,16639. .16676))
                /gene="B12N19.040"
                /note="similarity to heterokaryon incompatibility protein
                het-6, Neurospora crassa, TREMBL:AF208542_1"
                /codon_start=1
                /product="related to heterokaryon incompatibility protein
                het-6"
                /protein_id="CAD21065.1"
                /db_xref="GI:18376318"
                /db_xref="UniProt/TrEMBL:Q8X0I3"
                /translation="MEAHEGGERVGACNSGPVQTTGERYPYRQLDHENTIRLLFITPG
                SGEEASSTYSLLHAELGIIPYQALSYEWGLPSDDDPNITIDGHTVRIRKNLSEALKQI
                SSVIRYLDFLLLWIDALCINQNDDAEKSQQVREMGKIFSSAEQVLAWTGPMANDSDYA
                IDMLNAAPQDRIFDTQLPENYRARTAILAWCNRPYWKRIWIIQELFLAKRLVFMCGSR
                SISRRGNVVANCLWVILKTSEKSLPVVKSDNAEPWDTVLYRGKNNSYTLRDWLYVCAT
                SQIMASDTRDYIYAVLSISHETRSSLAQIIPDYTKSAETVFEDLNTLMMDERAREGCE
                SCLGKVKTFDYIRNKTINRWLIKTLGLNVQIGWHDESQYPSCNSCGRPEREYSKGELS
                AVKVRSEGKEFWTSEVHSLNFTAMVVENLWQGGG"
     exon       complement(15344. .16586)
                /gene="B12N19.040"
                /number=1
     intron     complement(16587. .16638)
                /gene="B12N19.040"
                /number=1
     exon       complement(16639. .16676)
                /gene="B12N19.040"
                /number=2
     gene       17463. .18879
                /gene="B12N19.050"
     gene       complement(join(17463. .18741,18857. .18879))
                /gene="B12N19.050"
     CDS        complement(join(17463. .18741,18857. .18879))
                /gene="B12N19.050"
                /note="similarity to heterokaryon incompatibility protein
                het-6, Neurospora crassa, TREMBL:AF208542_1"
                /codon_start=1
                /product="related to heterokaryon incompatibility protein
                het-6"
                /protein_id="CAD21066.1"
                /db_xref="GI:18376319"
                /db_xref="UniProt/TrEMBL:Q8X0I2"
                /translation="MPPLSYYTIGEDQGEDELPPAILPLTEEGGNAIRLLTLREPSSH
                EGETLSLEMHLSTTSEKYIALSYEWGQAEPDDPMILVNNRQVQIRKNLHEALRQISEH
                YSFYRKQRHDFKCPHCHGSVVDACKCGNRPEKDCHEKSLGPWTNLGDSSFWAKHDPPR
                LWIDALCINQNDNKEKSSQVGRMGTIFSSAKVVLSWIGLPRDGSNDALRMMRTLSMSY
                PELPEFVATGGLTDSLILSIASLCERTYWRRVWILQELYLAQKYVVMCGPESIASDSF
                DEALGQLLDLSLRPFLKSYNIMSRSAANAMVVSKRVSGLCSLVRWLIVVVNADFQATV
                PHDYIYAILAISSPGVIDMLKPGIDTTQIIVDYNKTVAEVYRQLLQSFEAGPSLMNHR
                RWLLKLALKMGLSQEEGVALLDSEHSKRCEVAKGSKSDLDY"
     exon       complement(17463. .18741)
                /gene="B12N19.050"
                /number=1
     intron     complement(18742. .18856)
                /gene="B12N19.050"
                /number=1
     exon       complement(18857. .18879)
                /gene="B12N19.050"
                /number=2
     gene       20541. .21062
                /gene="B12N19.060"
     gene       complement(20541. .21062)
                /gene="B12N19.060"
     CDS        complement(20541. .21062)
                /gene="B12N19.060"
                /note="weak similarity to proteinase umuD (EC 3.4.21.-) -
                Salmonella typhimurium, PIR:A36713"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="CAD21067.1"
                /db_xref="GI:18376320"
                /db_xref="UniProt/TrEMBL:Q8X0I1"
                /translation="MPTVMGLTTDSSSKTTTTLGNQDLKKPNGGSRLDTISKRNHHAA
                KEIETLFWRCLCDDPSAAPEYMAEDCILINPLLSSHPSTTPLSKDSEPNILEVFEKAE
                PWAGFRFHGDPLVVEIDLMAVALVYKVSLYTVDKKGKGTREVVASVSSSWRQTAGADW
                VLVAHHCQYVDDE"
     exon       complement(20541. .21062)
                /gene="B12N19.060"
```

```
                     /number=1
     repeat_region   21827. .21854
                     /note="28 bp poly(a) tandem repeat"
     gene            23273. .29225
                     /gene="B12N19.070"
     gene            complement(join(23273. .24898,24975. .28445,28506. .28670,
                     28743. .29225))
                     /gene="B12N19.070"
     CDS             complement(join(23273. .24898,24975. .28445,28506. .28670,
                     28743. .29225))
                     /gene="B12N19.070"
                     /note="similarity to regulator of PHR1, YER169w,
                     Saccharomyces cerevisiae, PIR:S50672"
                     /codon_start=1
                     /product="related to regulator of deoxyribodipyrimidine
                     photo-lyase PHR1"
                     /protein_id="CAD21068.1"
                     /db_xref="GI:18376321"

Alignment Scores:
Pred. No.:                15.7          Length:         42273
Score:                    74.00         Matches:        10
Percent Similarity:       100.00%       Conservative:   2
Best Local Similarity:    83.33%        Mismatches:     0
Query Match:              88.10%        Indels:         0
DB:                       15            Gaps:           0

US-09-858-332G-15 (1-12) x NCB12N19 (1-42273)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::||||||||||||||||||||:::||||||
Db      23041 CATAGCCATAACCATAACCATAACCATGACCATAAC 23076

RESULT 251
AC024772/c
LOCUS       AC024772               45870 bp    DNA     linear   INV 22-SEP-2004
DEFINITION  Caenorhabditis elegans cosmid Y40C5A, complete sequence.
ACCESSION   AC024772
VERSION     AC024772.1  GI:7140326
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 45870)
  AUTHORS   .
  CONSRTM   WormBase Consortium
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
  JOURNAL   Science 282 (5396), 2012-2018 (1998)
   PUBMED   9851916
REFERENCE   2  (bases 1 to 45870)
  AUTHORS   Kalicki,J. and Randall,J.
  TITLE     The sequence of C. elegans cosmid Y40C5A
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 45870)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 45870)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   5  (bases 1 to 45870)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2001) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   6  (bases 1 to 45870)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAY-2002) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   7  (bases 1 to 45870)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-AUG-2002) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   8  (bases 1 to 45870)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-NOV-2002) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   9  (bases 1 to 45870)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-APR-2003) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   10 (bases 1 to 45870)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-AUG-2003) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   11 (bases 1 to 45870)
  AUTHORS   .
  CONSRTM   WormBase Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-2004) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
COMMENT     Submitted by:
                  Genome Sequencing Center
                  Department of Genetics, Washington University
                  St. Louis , MO  63110, USA, and
                  Sanger Centre, Hinxton Hall
                  Cambridge CB10 IRQ, England
                  email: submissions@watson.wustl.edu and jes@sanger.ac.uk

            NOTICE:  This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate chemistry
            or covered by high quality data (i.e., phred quality >= 30); an
            attempt was made to resolve all sequencing problems, such as
            compressions and repeats; all regions were covered by sequence from
            more than one m13 subclone.

            For a graphical representation of this clone sequence and its
            analysis see:
            http://www.wormbase.org/db/seq/sequence?name=Y40C5A;class=Sequence

                        NEIGHBORING CLONE INFORMATION

            The 5' clone is F42C5, 400 bp overlap; the 3' clone is T12B3, 200
            bp overlap. Actual start of this clone is at base position 17680 of
            C06E4; actual end is at 5974 of F55G1.

            NOTES:

            Coding seqences below are the result of integration and manual
            review of the following data : computer analysis using the program
            Genefinder (P. Green and L. Hillier, personal communication), the
            large scale EST projects of Yuji Kohara
            (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C.
```

elegans ORFeome cloning project (http://worfdb.dfci.harvard.edu/), similarity to other proteins from BlastX analyses (http://blast.wustl.edu/), sequence conservation with C. briggsae using Jim Kent's WABA alignment program (Genome Research 10:1115-1125, 2000), individual C. elegans GenBank submissions, and personal communications with C. elegans researchers. tRNAs are predicted using the program tRNAscan-SE (Lowe, T.M. and Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

FEATURES Location/Qualifiers

source 1. .45870
/organism="Caenorhabditis elegans"
/mol_type="genomic DNA"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="Y40C5A"

gene complement(1532. .2606)
/gene="Y40C5A.1"
/locus_tag="Y40C5A.1"

CDS complement(join(1532. .1935,2192. .2296,2343. .2493, 2538. .2606))
/gene="Y40C5A.1"
/locus_tag="Y40C5A.1"
/standard_name="Y40C5A.1"
/note="contains similarity to Pfam domain PF00569 (Zinc finger, ZZ type); coded for by the following C. elegans cDNAs: CK580830, CK581299, CK582655, CK583705, CK585830, CK587547, OSTF133A11_1, OSTR133A11_1, yk1465f02.5, yk1698c08.3, yk1698c08.5"
/codon_start=1
/product="Hypothetical protein Y40C5A.1"
/protein_id="AAF60536.2"
/db_xref="GI:33621030"
/db_xref="WormBase:Y40C5A.1"
/translation="MSSIRYRRPPIIERRYERTPSYSTIPRDIHVSFYHNNYHRSKIL YLNKFNAMEKLIEAAHELYPSNGVYRLFSDSTIGAQNELNDSVQVMECVQNPDHILPH LHIRLEDYRPRSISRNYHEVNVPIIRTPSQHSIHQSVRFNDNKRHHHHHHHDSIRNFV HHDTPCRNCFRIIVGSRYHCLQCPDYDICGDCEKDLVHYEHALLRIVTPRVTSIPDYV TVNAPSYIFPEHRYDIKREHYHWF"

gene complement(3325. .6390)
/gene="ocr-4"
/locus_tag="Y40C5A.2"

CDS complement(join(3325. .3498,3546. .3791,3866. .4145, 4198. .4457,4499. .4664,4910. .5145,5192. .5293,5336. .5528, 5697. .6178,6220. .6390))
/gene="ocr-4"
/locus_tag="Y40C5A.2"
/standard_name="Y40C5A.2"
/note="contains similarity to Pfam domains PF00520 (Ion transport protein), PF00023 (Ankyrin repeat); coded for by the following C. elegans cDNAs: OSTF142H3_1, OSTR142H3_1, yk406b3.5"
/codon_start=1
/product="Osm-9 and capsaicin receptor-related protein 4"
/protein_id="AAF60537.1"
/db_xref="GI:7331849"
/db_xref="WormBase:Y40C5A.2"
/translation="MGNASSAVTAGIKNQVNASNNAINKLADMSGRGFLADLAMDAVR TGNHEELDKQILEKVQPMLYNNGQGEMIPLSDLIAQRHKERTGNTFAVPVPSTECRFI CWKLNSRGAVGETLLHTCFLAGLPNHMKLLAERLIAIFPKIINDFYLSDEYYGETVLH MGIVSENAEFVRYLLKSGADVHARCSGNFFTCDDQKGSRTDHPEVEHAIISKHTKYPG NFTMSSKKNPIFSHIYWGEYPLSFAACLSQPECIRMLVAHGADVNAADTNGNTILHIC TIHENYEMFKMALSLGADLHIQNRQNLTPLTLAAFLAKKEMMQRIVEEERTVNWTYGR TQSAAYPLEHVDSIEPTSGSINQKSVLTIAVYGEKVEHLSLLPHLLEQLVHHKWIAYG RKTLFAQLSMFCIYFICVTSCFLLRPSPFERQQHIANDLICFYDFKTRRFNESAVSTI IYHLLHLICVTGATLYLVQALLHIKNVGYHLYVLGLSGFPAKAIFLFSCILMVITFFL RIFCLDEAEDIVWIVIVLLTSLKFLFFCRGFKSVGPFVLMLYKIIIRDLMRFFLIYLV IVVGFSQAFYVIFLGYKRSDSGNREPSIMSNVAESYVRMFIMSLTEFTVFFEQLEECE HTVIGKITFVVYMLLVTLLLINMLIAMMTNTYTEVSGNSLEWLRQWSAIILMMEQSFD PATRLRYQRRYGIRFDGGEKLVLLLKDKMTEEEFECQRKKMHEERRKFREELRTNQRR RPVFICKQGANYFKSKHRHTNNKLIIPTG"

gene complement(6751. .16497)
/gene="Y40C5A.3"

CDS complement(join(6751. .6928,7068. .7329,7376. .7452, 7568. .7899,7978. .8248,8330. .8392,9207. .9356,9401. .9623, 9671. .10015,10168. .10206,10310. .10384,10430. .13399, 13489. .14619,14767. .14829,15138. .15625,15778. .15871, 15969. .16011,16098. .16195,16365. .16497))
/gene="Y40C5A.3"
/locus_tag="Y40C5A.3"
/standard_name="Y40C5A.3"
/note="contains similarity to Interpro domain IPR000194 (H+-transporting two-sector ATPase, alpha/beta subunit, central region); coded for by the following C. elegans cDNAs: yk305e9.5"
/codon_start=1
/product="Hypothetical protein Y40C5A.3"
/protein_id="AAF60538.1"

Alignment Scores:

Pred. No.:	17	Length:	45870
Score:	74.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	83.33%	Mismatches:	0
Query Match:	88.10%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-15 (1-12) x AC024772 (1-45870)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||:::||||||||||:::||||||||||||
Db       8036 CACAACCACAGCCACAACCACAGCCACAACCACAAC 8001
```

RESULT 252
CR382127_05
WPCOMMENT
Sequence split into 23 fragments LOCUS CR382127 Accession CR382127

Fragment Name	Begin	End
CR382127_00	1	110000
CR382127_01	100001	210000
CR382127_02	200001	310000
CR382127_03	300001	410000
CR382127_04	400001	510000
CR382127_05	500001	610000
CR382127_06	600001	710000
CR382127_07	700001	810000
CR382127_08	800001	910000
CR382127_09	900001	1010000
CR382127_10	1000001	1110000
CR382127_11	1100001	1210000
CR382127_12	1200001	1310000
CR382127_13	1300001	1410000
CR382127_14	1400001	1510000
CR382127_15	1500001	1610000
CR382127_16	1600001	1710000
CR382127_17	1700001	1810000
CR382127_18	1800001	1910000
CR382127_19	1900001	2010000
CR382127_20	2000001	2110000
CR382127_21	2100001	2210000
CR382127_22	2200001	2303261

Continuation (6 of 23) of CR382127 from base 500001 (CR382127 Yarrowia lipolytica chromos

Alignment Scores:

Pred. No.:	39.4	Length:	110000
Score:	74.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	83.33%	Mismatches:	0
Query Match:	88.10%	Indels:	0
DB:	15	Gaps:	0

US-09-858-332G-15 (1-12) x CR382127_05 (1-110000)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||||:::|||:::
```

```
Db      74733 CACAATCACAACCACAATCACAACCACAGTCACAGC 74768

RESULT 253
AC159703
LOCUS       AC159703               148802 bp    DNA     linear   INV 07-APR-2005
DEFINITION  Trypanosoma brucei chromosome 8 clone RPCI93-30K1, complete
            sequence.
ACCESSION   AC159703
VERSION     AC159703.1  GI:62359576
KEYWORDS    HTG.
SOURCE      Trypanosoma brucei
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 148802)
  AUTHORS   Ghedin,E., Blandin,G., Bartholomeu,D., Caler,E., Haas,B.,
            Hannick,L., Shallom,J., Hou,L., Djikeng,A., Feldblyum,T.,
            Hostetler,J., Johnson,J., Jones,K., Koo,H.L., Larkin,C., Pai,G.,
            Peterson,J., Khalak,H.G., Salzberg,S., Simpson,A.J., Tallon,L., Van
            Aken,S., Wanless,D., White,O., Wortman,J., Fraser,C.M. and
            El-Sayed,N.M.A.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 148802)
  AUTHORS   Haas,B., Blandin,G. and El-Sayed,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-APR-2005) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, nelsayed@tigr.org
COMMENT     BAC clone RPCI93-30K1 from Trypanosoma brucei chromosome 8.
            Putative protein coding regions (>30 codons) were predicted by
            combining the output of the gene finding algorithm GLIMMER with
            similarity data and manual annotation of open reading frames.
            Further refinement of gene models (additions, deletions and
            alterations to boundary coordinates) were done manually based on
            detailed comparisons with the T. cruzi and L. major genomes. Genes
            on BACs were assigned a systematic name based on the chromosome and
            the BAC from which they originated (e.g. Tb03.27F10.410). Gene
            products were assigned a putative function when they shared
            significant similarity with experimentally characterized gene
            products or when they contained functionally known protein domains.
            Gene products were labelled either 'hypothetical protein', or
            'hypothetical protein, conserved' in the case of products showing
            significant similarity with proteins or domains of unknown function
            from other organisms. Gene products of unknown function with
            predicted orthologs in Trypanosoma cruzi and Leishmania major
            genomes were automatically classified as conserved proteins. Short
            predicted coding sequences (less than 150 codons) with no
            additional evidence of coding potential were labelled 'hypothetical
            protein, unlikely'.
FEATURES             Location/Qualifiers
     source          1. .148802
                     /organism="Trypanosoma brucei"
                     /mol_type="genomic DNA"
                     /isolate="GUTat10.1"
                     /db_xref="taxon:5691"
                     /chromosome="8"
                     /clone="RPCI93-30K1"
     misc_feature    1. .21371
                     /note="annotation for this sequence region can be found on
                     overlapping clone 11J15."
     misc_feature    567. .587
                     /note="AT_rich"
     repeat_region   1052. .1074
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit="ta"
     misc_feature    3341. .3381
                     /note="AT_rich"
     misc_feature    3563. .3649
                     /note="T-rich"
     repeat_region   4450. .4475
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit="ta"
     repeat_region   4497. .4516
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit="ggaaa"
     misc_feature    7029. .7079
                     /note="GA-rich"
     repeat_region   12175. .12208
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit="ttg"
     repeat_region   13090. .13111
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit="ta"
     misc_feature    13233. .13256
                     /note="AT_rich"
     repeat_region   13356. .13375
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit="ttg"
     repeat_region   13417. .13461
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit="ta"
     repeat_region   13766. .13812
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit="caa"
     repeat_region   14376. .14409
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit="caa"
     repeat_region   14475. .14502
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit="ttttc"
     repeat_region   14627. .14654
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit="taa"
     misc_feature    14717. .14743
                     /note="AT_rich"
     repeat_region   14872. .14901
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit="gaaa"
     misc_feature    15099. .15127
                     /note="AT_rich"
     repeat_region   15448. .15492
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit="cca"
     repeat_region   15540. .15617
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit="gaatg"
     misc_feature    15738. .15762
                     /note="AT_rich"
     repeat_region   15985. .16007
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit="gaaa"
     misc_feature    16762. .16811
                     /note="CT-rich"
     repeat_region   16837. .16858
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit="ca"
     repeat_region   17325. .17381
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit="ttta"
     repeat_region   19235. .19254
```

```
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit="ta"
     gene            <21372. .>22322
                     /locus_tag="Tb927.8.6480"
                     /note="synonyms: Tb08.30K1.220, Tb08.11J15.05;
                     hypothetical protein, conserved"
     mRNA            <21372. .>22322
                     /locus_tag="Tb927.8.6480"
                     /product="hypothetical protein, conserved"
     CDS             <21372. .22322
                     /locus_tag="Tb927.8.6480"
                     /note="Should be merged with annotated gene Tb08.30K1.220
                     on overlapping BAC to construct the complete gene;
                     go_component: extracellular region [goid 0005576];
                     go_function: aminomethyltransferase activity [goid
                     0004047];
                     go_process: glycine decarboxylation via glycine cleavage
                     system [goid 0019464]"
                     /codon_start=1
                     /product="hypothetical protein, conserved"
                     /protein_id="AAX80054.1"
                     /db_xref="GI:62359621"
                     /translation="LQGLFTNDLRQLQPGGSLWGCFLHHTGRVMCDAYLYQSTRTPEG
                     QVTIMIDVHCGVADTLLEHLKEYRMRKKLEIRSAAEELVVVAAATIGNSISSCGDNAG
                     SSPSSSSATYGGDQELSGPQGVDSFDTLAETFTDPRSFALPATLRKMIVPRKGAPPTL
                     DSEKLYKKFLYAAGVGEGPEVFRPSKTLPFEANTDLLRGVSFHKGCYMGQELTHRTHV
                     MLVTRKRTVPLFLQGELFDGKGGEKTPHVEGTLVIGNQKVGEVLTACGNVGLGLLRLN
                     HVDITTRSFPGLSLSDGTTVDARIPEWWDEKELRKVLTKS"
     gene            <23296. .>26535
                     /locus_tag="Tb927.8.6490"
                     /note="synonym: Tb08.30K1.240; protein kinase, putative"
     mRNA            <23296. .>26535
                     /locus_tag="Tb927.8.6490"
                     /product="protein kinase, putative"
     CDS             23296. .26535
                     /locus_tag="Tb927.8.6490"
                     /EC_number="2.7.1.-"
                     /note="go_function: protein kinase activity [goid
                     0004672];
                     go_function: ATP binding [goid 0005524];
                     go_process: protein amino acid phosphorylation [goid
                     0006468]"
                     /codon_start=1
                     /product="protein kinase, putative"
                     /protein_id="AAX80010.1"
                     /db_xref="GI:62359577"
                     /translation="MGCGASSRVVESDDRERHKSPHSAAECSPKLGAAEVSPNAKLKE
                     VPSVSTSVIEGLKGHQGTFVMSKFSNDSMTTSASYLNQPNTNEALPPIQTAEASFAPS
                     DAADSTRGNGRSPFASNFTDSNVATSFLRIPHLNSEVDSDARFSSNAFESRCRSGSVV
                     STGVCNVPILPTFSHCNSRDSRFQSPCGVNRCDSYSPTARPHECSRRCGWGEYETLRR
                     LSINNSMDGTNGTIADTATTASSVEHSNNYAGHRYSFIGTDCGCSSLFYDPCERCSRC
                     ILDTEAVYRCDLCMLIFCTDCHAEIGAALHRHKLTSFRRAVNSNVTDSFLINKGRNRD
                     GNKVINEYVVIKVLGRGSHAKVNLVQHHQDRTFYAVKILRCDRTKKIHHGIVSKSSTA
                     SDDDLLREIAVMKFVSHPNLIKLKEVIDDVESHKVYVIMEYCAKGPVHVHGEPPLPPE
                     KVRKYGCDILSGLLQLHAQYLYHWDIKPANCLVDDNDVAKIADFGACGSSTRTCKVGG
                     TPAYSCPEQFAGTHIGGHVADSWAFAMTLYQMSHGTLAYSTTSISSLRDSMLDPTPLP
                     IQEGVEPELKDLLRRMLNKDMSQRMMLHDAMKHSYFSGYVTQSYTTVEPSNESAGCIT

Alignment Scores:
Pred. No.:              52.7          Length:         148802
Score:                  74.00         Matches:        10
Percent Similarity:     100.00%       Conservative:   2
Best Local Similarity:  83.33%        Mismatches:     0
Query Match:            88.10%        Indels:         0
DB:                     2             Gaps:           0

US-09-858-332G-15 (1-12) x AC159703 (1-148802)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::||||||||||||||||||||||||
Db      15452 CACCACCACCACCACAATCACAACCACAACCACAAC 15487
```

```
RESULT 254
CR626888/c
LOCUS       CR626888              156552 bp    DNA     linear   VRT 25-JUL-2005
DEFINITION  Zebrafish DNA sequence from clone CH211-250B7 in linkage group 16,
            complete sequence.
ACCESSION   CR626888
VERSION     CR626888.8  GI:71060553
KEYWORDS    HTG.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 156552)
  AUTHORS   Almeida,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUL-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests:
            http://www.sanger.ac.uk/Projects/D_rerio/faqs.shtml#dataeight
COMMENT     On Jul 21, 2005 this sequence version replaced gi:67514468.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep
            Clone-derived Zebrafish pUC subclones occasionally display
            inconsistency over the length of mononucleotide A/T runs and
            conserved TA repeats. Where this is found the longest good quality
            representation will be submitted.
            Any regions longer than 1kb tagged as misc-feature 'unsure' are
            part of a tandem repeat of more than 10kb in length where it has
            not been possible to anchor the base differences between repeat
            copies.  The region has been built up based on the repeat element
            to match the total size of repeat indicated by restriction digest,
            but repeat copies may not be in the correct order and the usual
            finishing criteria may not apply.
            CH211-250B7 is from a CHORI-211 BAC library
            VECTOR: pTARBAC2.1.
FEATURES             Location/Qualifiers
     source          1. .156552
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /chromosome="16"
                     /clone="CH211-250B7"
                     /clone_lib="CHORI-211"
ORIGIN

Alignment Scores:
Pred. No.:              55.3          Length:         156552
Score:                  74.00         Matches:        10
Percent Similarity:     100.00%       Conservative:   2
Best Local Similarity:  83.33%        Mismatches:     0
Query Match:            88.10%        Indels:         0
DB:                     5             Gaps:           0

US-09-858-332G-15 (1-12) x CR626888 (1-156552)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
```

```
                    |||||||||:::||||||||||||||||:::||||||
Db      40658 CATAATCATCATCATAATCATAATCATCATCATAAT 40623

RESULT 255
AC064874/c
LOCUS       AC064874               160655 bp    DNA     linear   PRI 30-APR-2005
DEFINITION  Homo sapiens BAC clone RP11-680O16 from 2, complete sequence.
ACCESSION   AC064874
VERSION     AC064874.4  GI:16799062
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 160655)
  AUTHORS   Doebber,A., Nguyen,C. and Haakenson,W.
  TITLE     The sequence of Homo sapiens BAC clone RP11-680O16
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 160655)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-APR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3  (bases 1 to 160655)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 160655)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JAN-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 160655)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 160655)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-APR-2005) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Nov 8, 2001 this sequence version replaced gi:15825638.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0680O16
            --------------.

            NOTICE:

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR:  pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-473L20; the clone sequenced
            to the right is RP11-485M18, 2000 bp overlap.  Actual start of this
            clone is at base position 1 of RP11-680O16; actual end is at base
            position 36903 of RP11-485M18.

            The sequence between 103323 and 103663 is covered only by a pcr
            product of clone DNA. Single stranded region exists between 103323
            and 103594. Data from AC073989 was used to finish this clone,
            AC064874.
FEATURES             Location/Qualifiers
     source          1. .160655
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="RP11-680O16"
                     /clone_lib="RPCI-11"
     gene            9699. .107851
                     /gene="CENTG2"
     mRNA            join(9699. .9757,18077. .18164,41483. .41568,45218. .45359,
                     50874. .51008,98279. .98406,99887. .100042,107759. .107851)
                     /gene="CENTG2"
     CDS             join(<9699. .9757,18077. .18164,41483. .41568,45218. .45359,
                     50874. .51008,98279. .98406,99887. .100042,107759. .>107851)
                     /gene="CENTG2"
                     /note="Homo sapiens centaurin, gamma 2 (CENTG2), mRNA.;
                     H_NH0680O16.1
                     This gene was based on gi(7662483)
                     Continued from H_NH0084G18.1
                     Continues as H_NH0485M18.1"
                     /codon_start=3
                     /product="unknown"
                     /protein_id="AAY24137.1"
                     /db_xref="GI:62988750"
                     /translation="AFVNSQEWTLSRSVPELKVGIVGNLASGKSALVHRYLTGTYVQE
                     ESPEGGRFKKEIVVDGQSYLLLIRDEGGPPEAQFAMWVDAVIFVFSLEDEISFQTVYH
                     YYSRMANYRNTSEIPLVLVGTQDAISSANPRVIDDARARKLSNDLKRCTYYETCATYG
                     LNVERVFQDVAQKIVATRKKQQLSIGPCKSLPNSPSHSSVCSAQVSAVHISQTSNGGG
                     SLSDYSSSVPSTPSTSQKELRIDVPPTANTPTPVRKQSKRRSNLFTSRKGSDPDKEKK
                     GLESRADSIGSGRAIPIKQ"
ORIGIN

Alignment Scores:
Pred. No.:                56.7          Length:          160655
Score:                    74.00         Matches:         10
Percent Similarity:       100.00%       Conservative:    2
Best Local Similarity:    83.33%        Mismatches:      0
Query Match:              88.10%        Indels:          0
DB:                       8             Gaps:            0

US-09-858-332G-15 (1-12) x AC064874 (1-160655)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||||||||:::||||||||||||||||:::||||||
Db     103380 CACAACCATCATCACAACCATAACCATCATCACAAC 103345

RESULT 256
AC159405/c
LOCUS       AC159405               164873 bp    DNA     linear   INV 07-APR-2005
DEFINITION  Trypanosoma brucei chromosome 8 clone RPCI93-11J15, complete
            sequence.
```

```
ACCESSION   AC159405
VERSION     AC159405.1  GI:62175120
KEYWORDS    HTG.
SOURCE      Trypanosoma brucei
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 164873)
  AUTHORS   Ghedin,E., Blandin,G., Bartholomeu,D., Caler,E., Haas,B.,
            Hannick,L., Shallom,J., Hou,L., Djikeng,A., Feldblyum,T.,
            Hostetler,J., Johnson,J., Jones,K., Koo,H.L., Larkin,C., Pai,G.,
            Peterson,J., Khalak,H.G., Salzberg,S., Simpson,A.J., Tallon,L., Van
            Aken,S., Wanless,D., White,O., Wortman,J., Fraser,C.M. and
            El-Sayed,N.M.A.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 164873)
  AUTHORS   Haas,B., Blandin,G. and El-Sayed,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-APR-2005) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, nelsayed@tigr.org
REFERENCE   3  (bases 1 to 164873)
  AUTHORS   Haas,B., Blandin,G. and El-Sayed,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-APR-2005) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, nelsayed@tigr.org
COMMENT     BAC clone RPCI93-11J15 from Trypanosoma brucei chromosome 8.
            Putative protein coding regions (>30 codons) were predicted by
            combining the output of the gene finding algorithm GLIMMER with
            similarity data and manual annotation of open reading frames.
            Further refinement of gene models (additions, deletions and
            alterations to boundary coordinates) were done manually based on
            detailed comparisons with the T. cruzi and L. major genomes. Genes
            on BACs were assigned a systematic name based on the chromosome and
            the BAC from which they originated (e.g. Tb03.27F10.410). Gene
            products were assigned a putative function when they shared
            significant similarity with experimentally characterized gene
            products or when they contained functionally known protein domains.
            Gene products were labelled either 'hypothetical protein', or
            'hypothetical protein, conserved' in the case of products showing
            significant similarity with proteins or domains of unknown function
            from other organisms. Gene products of unknown function with
            predicted orthologs in Trypanosoma cruzi and Leishmania major
            genomes were automatically classified as conserved proteins. Short
            predicted coding sequences (less than 150 codons) with no
            additional evidence of coding potential were labelled 'hypothetical
            protein, unlikely'.
FEATURES             Location/Qualifiers
     source          1. .164873
                     /organism="Trypanosoma brucei"
                     /mol_type="genomic DNA"
                     /isolate="GUTat10.1"
                     /db_xref="taxon:5691"
                     /chromosome="8"
                     /clone="RPCI93-11J15"
     gene            complement(<1. .>69)
                     /locus_tag="Tb927.8.6480"
                     /note="synonyms: Tb08.11J15.05, Tb08.30K1.220;
                     hypothetical protein, conserved"
     mRNA            complement(<1. .>69)
                     /locus_tag="Tb927.8.6480"
                     /product="hypothetical protein, conserved"
     CDS             complement(<1. .69)
                     /locus_tag="Tb927.8.6480"
                     /note="Should be merged with annotated gene Tb08.30K1.220
                     on overlapping BAC to construct the complete gene;
                     go_component: extracellular region [goid 0005576];
                     go_function: aminomethyltransferase activity [goid
                     0004047];
                     go_process: glycine decarboxylation via glycine cleavage
                     system [goid 0019464]"
                     /codon_start=1
                     /product="hypothetical protein, conserved"
                     /protein_id="AAX69333.1"
                     /db_xref="GI:62175185"
                     /translation="MGFKCLLSSRALLQVTGSVAHEF"
     gene            complement(<471. .>2108)
                     /locus_tag="Tb927.8.6470"
                     /note="synonym: Tb08.11J15.10; hypothetical protein,
                     conserved"
     mRNA            complement(<471. .>2108)
                     /locus_tag="Tb927.8.6470"
                     /product="hypothetical protein, conserved"
     CDS             complement(471. .2108)
                     /locus_tag="Tb927.8.6470"
                     /codon_start=1
                     /product="hypothetical protein, conserved"
                     /protein_id="AAX69271.1"
                     /db_xref="GI:62175123"
                     /translation="MPGCLEVRYIGESHGVGSYTLKPAKAGDVLLQEYPVAFAKAQRQ
                     DALEPHRICGGCGMILTSLKEECRRLALLALEATTGGSSQLDERVHGSSNACEEQKYE
                     HEAVGNRAGVTECAVEAPELAVIVSPDKLYDTFKDHEMWEHFQKGVEASDDIPEGGAD
                     SAEVRFCSMECKEHCLGGRGGRFVLPLLREPPCNELQLTSGVMEAQGIEDVDKLRTPS
                     ADTIMYQSTTLTAEVAVEVWPTKAHALATLEFLAEKYNARLRLVLFILARFLSDILST
                     GSVAANKQFEERLGERVQEFVQHYDEGASRGLSPEQRSFLRFSWKCVCRWITLCCVNE
                     GKVATAVAGSEDKFCEGLCSAERVCSSTAEWFPLQLYLRCFWVTDANAHMYVVVSPLY
                     SLLSERLPILSGIYRIQSETHFGKGRGENVCENEDSNVGSGREVDRLNRQMKILQFLF
                     QTVDPNRAHSRGVALYDAAAKINHSCAPSVRFVPTHGGVKAVVVALRDIPSGEEVRTS
                     YIEVGAYPTNKARREFLLSSYGFNCDCPLCVTECEPT"
     repeat_region   2118. .2137
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit="ta"
     gene            complement(<2589. .>3128)
                     /locus_tag="Tb927.8.6460"
                     /note="synonym: Tb08.11J15.30; hypothetical protein,
                     conserved"
     mRNA            complement(<2589. .>3128)
                     /locus_tag="Tb927.8.6460"
                     /product="hypothetical protein, conserved"
     CDS             complement(2589. .3128)
                     /locus_tag="Tb927.8.6460"
                     /codon_start=1
                     /product="hypothetical protein, conserved"
                     /protein_id="AAX69273.1"
                     /db_xref="GI:62175125"
                     /translation="MERIGEFTVMAFHNVRNAAALHDHFKKKKIDAAVIDAAYIVSHL
                     HLSVALHRLSTEPLITESGAKSVASSSTACGTSLVPERKVTARDVFSSLSHTRNLDRV
                     LERLTCASTTTAVVIVLHAPSSQHVEGVAAFVEGEPHRLSDNNRYCKEAAVRDFYGVG
                     EAETGLEAAVVNRLATSGI"
     repeat_region   3991. .4047
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit="taaa"
     gene            complement(<4091. .>4456)
                     /locus_tag="Tb927.8.6450"
                     /note="synonym: Tb08.11J15.50; inhibitor of cysteine
                     peptidase"
     mRNA            complement(<4091. .>4456)
                     /locus_tag="Tb927.8.6450"
                     /product="inhibitor of cysteine peptidase"
     CDS             complement(4091. .4456)
                     /locus_tag="Tb927.8.6450"
                     /note="99% similar to GB:CAD68976.1: inhibitor of cysteine
                     peptidase {Trypanosoma brucei} (PMID:12729889);
                     go_function: endopeptidase inhibitor activity [goid
                     0004866];
                     go_process: regulation of proteolysis and peptidolysis
                     [goid 0030162]"
                     /codon_start=1
                     /product="inhibitor of cysteine peptidase"
                     /protein_id="AAX69280.1"
                     /db_xref="GI:62175132"
                     /translation="MSHNLFTEEDNNKTIRMVIGETFTIELKSNPTTGYTWLRSGLAG
                     TELSDCTFAIQSKFNNRAPHDNHKNHRRLLVGAGGTMVLEVKALKPGKHTLSLAYGRP
                     WVGFNAAAKRYNIHVEATA"
     repeat_region   4514. .4535
```

```
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit="tg"
     misc_feature    4561. .4610
                     /note="GA-rich"
     repeat_region   5365. .5387
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit="tttc"
     misc_feature    5610. .5634
                     /note="AT_rich"
     repeat_region   5755. .5832
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit="cattc"
     repeat_region   5880. .5924
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit="tgg"
     misc_feature    6245. .6273
                     /note="AT_rich"
     repeat_region   6471. .6500
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit="tttc"
     misc_feature    6629. .6655
                     /note="AT_rich"
     repeat_region   6718. .6745
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit="tta"
     repeat_region   6870. .6897
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit="gaaaa"
     repeat_region   6963. .6996
                     /note="microsatellite"
                     /rpt_type=tandem
                     /rpt_unit="ttg"
     gene            complement(<7125. .>7679)
                     /locus_tag="Tb927.8.6440"
                     /note="synonym: Tb08.11J15.100; RNA-binding protein,
                     putative"
     mRNA            complement(<7125. .>7679)
                     /locus_tag="Tb927.8.6440"
                     /product="RNA-binding protein, putative"
     CDS             complement(7125. .7679)
                     /locus_tag="Tb927.8.6440"
                     /note="go_function: RNA binding [goid 0003723];
                     go_process: RNA processing [goid 0006396]"
                     /codon_start=1

Alignment Scores:
Pred. No.:              58.1          Length:          164873
Score:                  74.00         Matches:         10
Percent Similarity:     100.00%       Conservative:    2
Best Local Similarity:  83.33%        Mismatches:      0
Query Match:            88.10%        Indels:          0
DB:                     2             Gaps:            0

US-09-858-332G-15 (1-12) x AC159405 (1-164873)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::|||||||||||||||||||||||||
Db       5920 CACCACCACCACCACAATCACAACCACAACCACAAC 5885

RESULT 257
AC119200
LOCUS       AC119200              181206 bp    DNA     linear   ROD 17-AUG-2004
DEFINITION  Mus musculus chromosome 8, clone RP24-181K12, complete sequence.
ACCESSION   AC119200
VERSION     AC119200.5  GI:51315557
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 181206)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 8, clone RP24-181K12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 181206)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 181206)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
            Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J.,
            Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
            DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
            Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,
            Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,
            Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
            MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
            Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
            Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-2004) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 181206)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
            Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J.,
            Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
            DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
            Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,
            Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,
            Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
            MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
```

```
            Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
            Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
            Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-AUG-2004) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Aug 17, 2004 this sequence version replaced gi:45332560.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@broad.mit.edu
            -------------- Project Information
                Center project name: L25187
                Center clone name: 181_K_12
            --------------.
FEATURES             Location/Qualifiers
     source          1. .181206
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="8"
                     /map="8"
                     /clone="RP24-181K12"
                     /clone_lib="RPCI-24 Male Mouse BAC"
     misc_feature    1. .4
                     /note="clone_boundary
                     clone_end:SP6
                     site:MboI"
     repeat_region   326. .364
                     /rpt_family="(CA)n"
     repeat_region   complement(508. .584)
                     /rpt_family="ID4"
     repeat_region   complement(607. .823)
                     /rpt_family="B3A"
     repeat_region   1055. .1127
                     /rpt_family="(CATATA)n"
     repeat_region   complement(1183. .1291)
                     /rpt_family="PB1D10"
     repeat_region   1513. .1558
                     /rpt_family="(TTTG)n"
     repeat_region   complement(1559. .1694)
                     /rpt_family="B1_MM"
     repeat_region   complement(4038. .4190)
                     /rpt_family="B1F"
     repeat_region   4274. .4296
                     /rpt_family="AT_rich"
     repeat_region   complement(4509. .4657)
                     /rpt_family="URR1A"
     repeat_region   complement(5088. .5201)
                     /rpt_family="PB1D10"
     repeat_region   5555. .5596
                     /rpt_family="(TTGG)n"
     repeat_region   complement(5597. .5667)
                     /rpt_family="B1_MM"
     repeat_region   complement(6479. .6661)
                     /rpt_family="B2_Mm2"
     repeat_region   complement(6859. .7071)
                     /rpt_family="B3"
     repeat_region   7133. .7222
                     /rpt_family="B1_MM"
     repeat_region   complement(7396. .7495)
                     /rpt_family="B1_MM"
     repeat_region   complement(7944. .8240)
                     /rpt_family="B4"
     repeat_region   complement(8450. .8755)
                     /rpt_family="MLT1A"
     repeat_region   complement(9020. .9194)
                     /rpt_family="B3"
     repeat_region   10669. .10716
                     /rpt_family="AT_rich"
     repeat_region   10733. .10813
                     /rpt_family="ID_B1"
     repeat_region   11485. .11637
                     /rpt_family="Lx6"
     repeat_region   12470. .12779
                     /rpt_family="ORR1D"
     repeat_region   complement(12890. .12976)
                     /rpt_family="MTE"
     repeat_region   15828. .15929
                     /rpt_family="PB1D10"
     repeat_region   complement(16978. .17053)
                     /rpt_family="B1_MM"
     repeat_region   17531. .17599
                     /rpt_family="ID3"
     repeat_region   complement(19351. .19727)
                     /rpt_family="ORR1C"
     repeat_region   complement(19793. .19854)
                     /rpt_family="ID5"
     repeat_region   19988. .20178
                     /rpt_family="B3"
     repeat_region   20777. .20800
                     /rpt_family="(TTTTG)n"
     repeat_region   21310. .21330
                     /rpt_family="AT_rich"
     repeat_region   22109. .22129
                     /rpt_family="AT_rich"
     repeat_region   22478. .22614
                     /rpt_family="B3"
     repeat_region   22616. .22691
                     /rpt_family="(TC)n"
     repeat_region   24036. .24167
                     /rpt_family="B1F"
     repeat_region   24275. .24417
                     /rpt_family="B3"
     repeat_region   24418. .24439
                     /rpt_family="AT_rich"
     repeat_region   complement(26756. .26967)
                     /rpt_family="MLT1J"
     repeat_region   27115. .27274
                     /rpt_family="ID_B1"
     repeat_region   complement(27284. .27369)
                     /rpt_family="B4A"

Alignment Scores:
Pred. No.:                63.7          Length:          181206
Score:                    74.00         Matches:         10
Percent Similarity:       100.00%       Conservative:    2
Best Local Similarity:    83.33%        Mismatches:      0
Query Match:              88.10%        Indels:          0
DB:                       9             Gaps:            0

US-09-858-332G-15 (1-12) x AC119200 (1-181206)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||:::||||||||||||||||||||||||
Db      47253 CATCACCACCACCACAACCACAATCACAACCACAAC 47288

RESULT 258
AC133201
LOCUS       AC133201               184783 bp    DNA     linear   ROD 25-NOV-2003
DEFINITION  Mus musculus BAC clone RP23-74A8 from chromosome 3, complete
            sequence.
ACCESSION   AC133201
VERSION     AC133201.2  GI:32964984
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
```

```
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 184783)
  AUTHORS   Abbott,S. and Haglund,K.
  TITLE     The sequence of Mus musculus BAC clone RP23-74A8
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 184783)
  AUTHORS   Wilson,R.
  TITLE     Sequencing of Mus musculus
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 184783)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 184783)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUL-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   5  (bases 1 to 184783)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-NOV-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Jul 18, 2003 this sequence version replaced gi:22758561.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: M_BA0074A08
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-23 BAC Library has been constructed by Kazutoyo Osegawa
            and Minako Tateno in the laboratory of Pieter de Jong
            (http://www.chori.org) from female C57BL/6J mouse kidney and/or
            brain genomic DNA.  The clone and detailed information can be
            obtained from Research Genetics, Inc. (http://www.resgen.com) or
            Pieter de Jong and coworkers at http://www.chori.org

            NEIGHBORING SEQUENCE INFORMATION:
            This sequence is the entire insert of the clone. This clone is
            overlapped by AC125109.
FEATURES              Location/Qualifiers
     source           1. .184783
                      /organism="Mus musculus"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:10090"
                      /chromosome="3"
                      /map="3"
                      /clone="RP23-74A8"
                      /clone_lib="RPCI-23"
     repeat_region    14. .1522
                      /rpt_family="L1"
     repeat_region    4402. .4685
                      /rpt_family="ERVK"
     repeat_region    4655. .4768
                      /rpt_family="ERVK"
     repeat_region    8489. .8803
                      /rpt_family="L1"
     repeat_region    9181. .9295
                      /rpt_family="MaLR"
     repeat_region    11158. .11915
                      /rpt_family="L1"
     repeat_region    13064. .13214
                      /rpt_family="B4"
     repeat_region    16240. .17038
                      /rpt_family="L1"
     repeat_region    17957. .18092
                      /rpt_family="B4"
     repeat_region    19264. .19550
                      /rpt_family="MaLR"
     repeat_region    19880. .20065
                      /rpt_family="L2"
     repeat_region    20287. .20522
                      /rpt_family="RMER15"
     repeat_region    20924. .21061
                      /rpt_family="MaLR"
     repeat_region    23426. .23625
                      /rpt_family="MER1_type"
     repeat_region    25928. .26112
                      /rpt_family="B2"
     repeat_region    26166. .26274
                      /rpt_family="B4"
     repeat_region    27548. .27683
                      /rpt_family="L1"
     repeat_region    27766. .28202
                      /rpt_family="L1"
     repeat_region    28593. .28966
                      /rpt_family="ERVK"
     repeat_region    28984. .29180
                      /rpt_family="L1"
     repeat_region    29181. .29259
                      /rpt_family="ERVK"
     repeat_region    29268. .30207
                      /rpt_family="L1"
     repeat_region    31292. .31715
                      /rpt_family="L1"
     repeat_region    31709. .31758
                      /rpt_family="L1"
     repeat_region    32132. .32258
                      /rpt_family="MaLR"
     repeat_region    32233. .32452
                      /rpt_family="MaLR"
     repeat_region    35186. .35394
                      /rpt_family="B2"
     repeat_region    42806. .42937
                      /rpt_family="Alu"
     repeat_region    43573. .43946
                      /rpt_family="MaLR"
     repeat_region    45582. .45926
                      /rpt_family="MaLR"
     repeat_region    49112. .49310
                      /rpt_family="B2"
     repeat_region    49594. .50417
                      /rpt_family="RMER13A"
     repeat_region    51429. .52052
                      /rpt_family="L1"
     repeat_region    52656. .53008
                      /rpt_family="MaLR"
     repeat_region    53580. .53645
                      /rpt_family="ID"
```

```
     repeat_region   53813. .53865
                     /rpt_family="ERV1"
     repeat_region   54072. .54417
                     /rpt_family="MaLR"
     repeat_region   54829. .55619
                     /rpt_family="L1"
     repeat_region   55619. .56760
                     /rpt_family="L1"
     repeat_region   57099. .57505
                     /rpt_family="L1"
     repeat_region   57509. .58049
                     /rpt_family="L1"
     repeat_region   59024. .59097
                     /rpt_family="ERV1"
     repeat_region   59253. .59443
                     /rpt_family="B4"
     repeat_region   61544. .61696
                     /rpt_family="L1"
     repeat_region   62234. .62421
                     /rpt_family="L1"
     repeat_region   62970. .63313
                     /rpt_family="L1"
     repeat_region   63315. .63482
                     /rpt_family="L1"
     repeat_region   63522. .63992
                     /rpt_family="L1"
     repeat_region   64733. .64908
                     /rpt_family="L1"
     repeat_region   66709. .66848
                     /rpt_family="Alu"
     repeat_region   67117. .67236
                     /rpt_family="MaLR"
     repeat_region   67358. .67627
                     /rpt_family="L1"
     repeat_region   67782. .67867
                     /rpt_family="L1"
     unsure          68309. .68547
                     /note="Unresolved simple sequence repeat."
     repeat_region   69521. .69808
                     /rpt_family="B4"
     repeat_region   72400. .72540
                     /rpt_family="L1"
     repeat_region   72528. .73282
                     /rpt_family="L1"
     repeat_region   73898. .74023
                     /rpt_family="B4"
     repeat_region   74654. .75270

Alignment Scores:
Pred. No.:              64.9          Length:       184783
Score:                  74.00         Matches:      10
Percent Similarity:     100.00%       Conservative: 2
Best Local Similarity:  83.33%        Mismatches:   0
Query Match:            88.10%        Indels:       0
DB:                     9             Gaps:         0

US-09-858-332G-15 (1-12) x AC133201 (1-184783)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||:::||||||||||||||:::||||||
Db      35919 CACAACCACCACCACAACCACAACCACCACCACAAC 35954

RESULT 259
AL683804
LOCUS       AL683804              198631 bp    DNA     linear   ROD 09-FEB-2005
DEFINITION  Mouse DNA sequence from clone RP23-17F2 on chromosome 1 Contains  a
            SEC61 gamma subunit (Sec61g) pseudogene, a limb expression 1
            homolog (chicken) (Lix1) pseudogene, a novel pseudogene, the 5' end
            of the gene for an ortholog of human Ras association (RalGDS/AF-6)
            and pleckstrin homology domains 1 RAPH1 and a CpG island, complete
            sequence.
ACCESSION   AL683804
VERSION     AL683804.15  GI:24366485
KEYWORDS    HTG; CpG island; Lix1; RalGDS/AF-6; RAPH1; SEC61.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 198631)
  AUTHORS   Skuce,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-FEB-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 24, 2002 this sequence version replaced gi:24210306.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep --------------
            Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
            Sequence from the Mouse Genome Sequencing Consortium whole genome
            shotgun may have been used to confirm this sequence.  Sequence data
            from the whole genome shotgun alone has only been used where it has
            a phred quality of at least 30.
            RP23-17F2 is from the RPCI-23 Mouse BAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .198631
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="1"
                     /clone="RP23-17F2"
                     /clone_lib="RPCI-23"
     gene            complement(join(1. .533,3127. .3269,5004. .5124,5814. .5912,
                     8013. .8123,9069. .9212,10144. .10209,11345. .11466,
                     12414. .12573,13107. .13184,20006. .20080,28569. .28649,
                     35181. .35689,36973. .37078,42876. .42995,76154. .76358))
                     /gene="RP23-17F2.1"
                     /locus_tag="RP23-17F2.1-003"
     mRNA            complement(join(1. .533,3127. .3269,5004. .5124,5814. .5912,
                     8013. .8123,9069. .9212,10144. .10209,11345. .11466,
                     12414. .12573,13107. .13184,20006. .20080,28569. .28649,
                     35181. .35689,36973. .37078,42876. .42995,76154. .76358))
                     /gene="RP23-17F2.1"
                     /locus_tag="RP23-17F2.1-003"
                     /product="ortholog of human Ras association (RalGDS\/AF-6)
                     and pleckstrin homology domains 1 RAPH1"
                     /note="match: cDNAs: Em:AY523978.1"
     CDS             complement(join(531. .533,3127. .3269,5004. .5124,5814. .5912,
                     8013. .8123,9069. .9212,10144. .10209,11345. .11466,
                     12414. .12573,13107. .13184,20006. .20080,28569. .28649,
                     35181. .35689,36973. .37078,42876. .42995))
                     /gene="RP23-17F2.1"
                     /locus_tag="RP23-17F2.1-003"
                     /standard_name="OTTMUSP00000002152"
                     /note="match: proteins: Tr:AAS16936"
                     /codon_start=1
                     /product="ortholog of human Ras association (RalGDS\/AF-6)
                     and pleckstrin homology domains 1 RAPH1"
```

/protein_id="CAI24281.1"
/db_xref="GI:56206998"
/translation="MEQLSDEEIDHGAEEDSDKEDQDLDKMFGAWLGELDRLTQSLDS
DKPKEPVKRSPLRQETNMANFSYRFSIYNLNEALNQGETVDLDALMADLCSIEQELSS
IGSGNSKRQVTEPKVPQKLPAGRHTSKHGTLRGPSSSSNRTTKPSHANYSLDDITAQL
EQASLSMDEAAQQSLVEDPKPLVTNQHRRTASAGTVSDAEARSISNSSRSSITSAASS
MDSLDIDKVTRPQELDLTTHQGQPITEHAISLSCPSKQAKHHSDFTEEQAELTPHSYL
DRETSLLLRNIAGKPSHLLTKEEQAAKLKAEKIRVALEKIKEAQVKKLVIRVHMSDDS
SKTMMVDERQTVRQVLDNLMDKSHCGYSLDWSLVETISELQMERIFEDHENLVENLLN
WTRDSQNKLIFMERIEKYALFKNPQNYLLGKKETAEMADRNKEVLLEECFCGSSVTVP
EIEGVLWLKDDGKKSWKKRYFLLRASGIYYVPKGKAKVSRDLVCFLQLDHVNVYYGQD
YRNKYKAPTDCCLALKHPQIQKKSQYIKYLCCDDARTLHQWVNGIRIAKYGKQLYTNY
QEALKRTESAYDWTSLSSSSIKSGSSSSSIPESQSNHSNQSDSGVSDTLPTGHIRSQS
TVSSIFSEAWKRGTQLEESSK"
gene complement(join(<2912. .3269,5004. .5124,5814. .5912,
8013. .8123,9069. .9212,10144. .10209,11345. .11466,
12414. .12573,13107. .13184,20006. .20080,28569. .28649,
35181. .35689,36973. .37078,42876. .42995,76154. .76358))
/gene="RP23-17F2.1"
/locus_tag="RP23-17F2.1-002"
mRNA complement(join(<2912. .3269,5004. .5124,5814. .5912,
8013. .8123,9069. .9212,10144. .10209,11345. .11466,
12414. .12573,13107. .13184,20006. .20080,28569. .28649,
35181. .35689,36973. .37078,42876. .42995,76154. .76358))
/gene="RP23-17F2.1"
/locus_tag="RP23-17F2.1-002"
/product="ortholog of human Ras association (RalGDS\/AF-6)
and pleckstrin homology domains 1 RAPH1"
/note="match: ESTs: Em:AI666558.1 Em:AI847203.1
Em:AW494129.1 Em:BI134311.1 Em:BQ829475.1 Em:BQ830930.1
Em:BY372951.1 Em:BY739987.1 Em:CA534431.1 Em:CK345747.1
match: cDNAs: Em:AB053311.1 Em:AY523977.1"
CDS complement(join(<2912. .3269,5004. .5124,5814. .5912,
8013. .8123,9069. .9212,10144. .10209,11345. .11466,
12414. .12573,13107. .13184,20006. .20080,28569. .28649,
35181. .35689,36973. .37078,42876. .42995))
/gene="RP23-17F2.1"
/locus_tag="RP23-17F2.1-002"
/standard_name="OTTMUSP00000002122"
/note="match: proteins: Tr:AAS16935 Tr:Q96Q37"
/codon_start=1
/product="ortholog of human Ras association (RalGDS\/AF-6)
and pleckstrin homology domains 1 RAPH1"
/protein_id="CAI24282.1"
/db_xref="GI:56206999"
/translation="MEQLSDEEIDHGAEEDSDKEDQDLDKMFGAWLGELDRLTQSLDS
DKPKEPVKRSPLRQETNMANFSYRFSIYNLNEALNQGETVDLDALMADLCSIEQELSS
IGSGNSKRQVTEPKVPQKLPAGRHTSKHGTLRGPSSSSNRTTKPSHANYSLDDITAQL
EQASLSMDEAAQQSLVEDPKPLVTNQHRRTASAGTVSDAEARSISNSSRSSITSAASS
MDSLDIDKVTRPQELDLTTHQGQPITEHAISLSCPSKQAKHHSDFTEEQAELTPHSYL
DRETSLLLRNIAGKPSHLLTKEEQAAKLKAEKIRVALEKIKEAQVKKLVIRVHMSDDS
SKTMMVDERQTVRQVLDNLMDKSHCGYSLDWSLVETISELQMERIFEDHENLVENLLN
WTRDSQNKLIFMERIEKYALFKNPQNYLLGKKETAEMADRNKEVLLEECFCGSSVTVP
EIEGVLWLKDDGKKSWKKRYFLLRASGIYYVPKGKAKVSRDLVCFLQLDHVNVYYGQD
YRNKYKAPTDCCLALKHPQIQKKSQYIKYLCCDDARTLHQWVNGIRIAKYGKQLYTNY
QEALKRTESAYDWTSLSSSSIKSGSSSSSIPESQSNHSNQSDSGVSDTLPTGHIRSQS
TVSSIFSEAWKRGTQLEESSKVTHKHYFTYKSVSSSVQKPVPFIPLLQSHLVPFHWSF
LGETILALSTVTCASLRSSFFMSVIYTLLLLPPP"
gene complement(join(<3127. .3269,5004. .5124,5814. .5912,
8013. .8123,9069. .9212,10144. .10209,11345. .11466,
12414. .12573,13107. .13184,35181. .35689,36973. .37078,
42876. .42995,76154. .76358))
/gene="RP23-17F2.1"
/locus_tag="RP23-17F2.1-001"
mRNA complement(join(<3127. .3269,5004. .5124,5814. .5912,
8013. .8123,9069. .9212,10144. .10209,11345. .11466,
12414. .12573,13107. .13184,35181. .35689,36973. .37078,
42876. .42995,76154. .76358))
/gene="RP23-17F2.1"
/locus_tag="RP23-17F2.1-001"
/product="ortholog of human Ras association (RalGDS\/AF-6)
and pleckstrin homology domains 1 RAPH1"
/note="match: ESTs: Em:CB246004.1
match: cDNAs: Em:AB051468.1 Em:AJ584699.1"
CDS complement(join(<3127. .3269,5004. .5124,5814. .5912,
8013. .8123,9069. .9212,10144. .10209,11345. .11466,
12414. .12573,13107. .13184,35181. .35689,36973. .37078,
42876. .42995))
/gene="RP23-17F2.1"
/locus_tag="RP23-17F2.1-001"
/standard_name="OTTMUSP00000002151"
/note="match: proteins: Tr:CAE48361 Tr:Q9C0I2"
/codon_start=1
/product="ortholog of human Ras association (RalGDS\/AF-6)
and pleckstrin homology domains 1 RAPH1"
/protein_id="CAI24283.1"
/db_xref="GI:56207000"
/translation="MEQLSDEEIDHGAEEDSDKEDQDLDKMFGAWLGELDRLTQSLDS
DKPKEPVKRSPLRQETNMANFSYRFSIYNLNEALNQGETVDLDALMADLCSIEQELSS
IGSGNSKRQVTEPKVPQKLPAGRHTSKHGTLRGPSSSSNRTTKPSHANYSLDDITAQL
EQASLSMDEAAQQSLVEDPKPLVTNQHRRTASAGTVSDAEARSISNSSRSSITSAASS
MDSLDIDKVTRPQELDLTTHQGQPITEEEQAAKLKAEKIRVALEKIKEAQVKKLVIRV
HMSDDSSKTMMVDERQTVRQVLDNLMDKSHCGYSLDWSLVETISELQMERIFEDHENL
VENLLNWTRDSQNKLIFMERIEKYALFKNPQNYLLGKKETAEMADRNKEVLLEECFCG
SSVTVPEIEGVLWLKDDGKKSWKKRYFLLRASGIYYVPKGKAKVSRDLVCFLQLDHVN
VYYGQDYRNKYKAPTDCCLALKHPQIQKKSQYIKYLCCDDARTLHQWVNGIRIAKYGK
QLYTNYQEALKRTESAYDWTSLSSSSIKSGSSSSSIPESQSNHSNQSDSGVSDTLPTG
HIRSQSTVSSIFSEAWKRGTQLEESSK"
gene complement(32428. .32655)
/locus_tag="RP23-17F2.2-001"
/pseudo
CDS complement(32428. .32655)
/locus_tag="RP23-17F2.2-001"
/note="match: proteins: Sw:P60058 Sw:P60059 Sw:P60060
Sw:Q09827 Sw:Q7SZU9 Sw:Q7T207 Sw:Q7Z1B8 Sw:Q8I7D9
Sw:Q962X7 Sw:Q9C2D4 Sw:Q9V668 Sw:Q9VWE9 Tr:AAQ96232
Tr:AAQ96234 Tr:AAQ96238 Tr:O96183 Tr:Q7R7E5 Tr:Q7TQ10
Tr:Q9V932"
/pseudo
/codon_start=1
/product="SEC61, gamma subunit (Sec61g) pseudogene"
gene complement(join(<35277. .35689,36973. .37078,42876. .42995,
76643. .76693))
/gene="RP23-17F2.1"
/locus_tag="RP23-17F2.1-004"
mRNA complement(join(<35277. .35689,36973. .37078,42876. .42995,
76643. .76693))
/gene="RP23-17F2.1"

Alignment Scores:

Pred. No.:	69.6	Length:	198631
Score:	74.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	83.33%	Mismatches:	0
Query Match:	88.10%	Indels:	0
DB:	9	Gaps:	0

US-09-858-332G-15 (1-12) x AL683804 (1-198631)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||:::|||:::||||||
Db     172467 CACAACCACAACCACAACCACCACCACCACCACAAC 172502
```

RESULT 260
AC006755/c
LOCUS AC006755 199917 bp DNA linear HTG 23-FEB-1999
DEFINITION Caenorhabditis elegans clone Y40C5, *** SEQUENCING IN PROGRESS ***.
ACCESSION AC006755
VERSION AC006755.1 GI:4263183
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 199917)
AUTHORS Waterston,R.H.

```
  TITLE     The sequence of Caehorhabditis elegans clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 199917)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   199917: contig of 199917 bp in length.
FEATURES             Location/Qualifiers
     source          1. .199917
                     /organism="Caenorhabditis elegans"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:6239"
                     /clone="Y40C5"
ORIGIN

Alignment Scores:
Pred. No.:                70              Length:         199917
Score:                    74.00           Matches:        10
Percent Similarity:       100.00%         Conservative:   2
Best Local Similarity:    83.33%          Mismatches:     0
Query Match:              88.10%          Indels:         0
DB:                       14              Gaps:           0

US-09-858-332G-15 (1-12) x AC006755 (1-199917)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||||||||:::|||||||||:::||||||||||||
Db      62993 CACAACCACAGCCACAACCACAGCCACAACCACAAC 62958

RESULT 261
AC101656
LOCUS       AC101656               209398 bp    DNA     linear   HTG 06-MAY-2004
DEFINITION  Mus musculus chromosome 1 clone RP23-306M24 map 1, *** SEQUENCING
            IN PROGRESS ***.
ACCESSION   AC101656
VERSION     AC101656.6  GI:47059725
KEYWORDS    HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 209398)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 1, clone RP23-306M24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 209398)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 209398)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
            Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J.,
            Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
            DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
            Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,
            Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,
            Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
            MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
            Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
            Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-2004) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 6, 2004 this sequence version replaced gi:47026541.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@broad.mit.edu
            -------------- Project Information
                Center project name: L17152
                Center clone name: 306_M_24
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   209398: contig of 209398 bp in length.
FEATURES             Location/Qualifiers
     source          1. .209398
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="1"
                     /map="1"
                     /clone="RP23-306M24"
                     /clone_lib="RPCI-23 Female Mouse BAC"
ORIGIN

Alignment Scores:
Pred. No.:                73.2            Length:         209398
Score:                    74.00           Matches:        10
Percent Similarity:       100.00%         Conservative:   2
Best Local Similarity:    83.33%          Mismatches:     0
```

```
Query Match:            88.10%      Indels:        0
DB:                     14          Gaps:          0

US-09-858-332G-15 (1-12) x AC101656 (1-209398)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||:::|||:::||||||
Db      33749 CACAACCACAACCACAACCACCACCACCACAAC 33784

RESULT 262
AC148987/c
LOCUS       AC148987               222379 bp    DNA     linear   ROD 28-JAN-2005
DEFINITION  Mus musculus BAC clone RP24-322F4 from 6, complete sequence.
ACCESSION   AC148987
VERSION     AC148987.3  GI:57012494
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 222379)
  AUTHORS   Kruchowski,S., Cotton,M. and Haglund,K.
  TITLE     The sequence of Mus musculus BAC clone RP24-322F4
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 222379)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-APR-2004) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 222379)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-2004) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 222379)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JAN-2005) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   5  (bases 1 to 222379)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JAN-2005) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jan 4, 2005 this sequence version replaced gi:52077884.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: M_BB0322F04
            --------------.

            NOTICE:

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-24 BAC Library has been constructed by Pieter de Jong and
            coworkers (http://www.chori.org) from male C57BL/6J mouse spleen
            and/or brain genomic DNA. The clone and detailed information can be
            obtained from Pieter de Jong and coworkers at http://www.chori.org

            NEIGHBORING SEQUENCE INFORMATION:
            This sequence is the entire insert of the clone.
FEATURES                Location/Qualifiers
     source             1. .222379
                        /organism="Mus musculus"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:10090"
                        /chromosome="6"
                        /map="6"
                        /clone="RP24-322F4"
                        /clone_lib="RPCI-24"
     misc_feature       155952. .155971
                        /note="Sequence derived from PCR product of project DNA."
     unsure             188554. .188664
                        /note="Sequence derived from one plasmid subclone."
     unsure             191189. .191395
                        /note="Unresolved simple sequence repeat."
ORIGIN

Alignment Scores:
Pred. No.:              77.6            Length:         222379
Score:                  74.00           Matches:        10
Percent Similarity:     100.00%         Conservative:   2
Best Local Similarity:  83.33%          Mismatches:     0
Query Match:            88.10%          Indels:         0
DB:                     9               Gaps:           0

US-09-858-332G-15 (1-12) x AC148987 (1-222379)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||||||:::|||:::
Db      85176 CACAACCACAACCACAACCACAACCACCACCACCAC 85141

RESULT 263
AC103448/c
LOCUS       AC103448               222776 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-201P18, *** SEQUENCING IN PROGRESS
            ***.
ACCESSION   AC103448
VERSION     AC103448.6  GI:30521383
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 222776)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
```

```
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 222776)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-NOV-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 222776)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:24953284.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GHPA
                Center clone name: CH230-201P18
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 192937 bases at least Q40
                Consensus quality: 196426 bases at least Q30
                Consensus quality: 198277 bases at least Q20
                Estimated insert size: 205228; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   222776: contig of 222776 bp in length.
FEATURES            Location/Qualifiers
     source         1. .222776
                    /organism="Rattus norvegicus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10116"
                    /clone="CH230-201P18"
     misc_feature   176922. .179176
                    /note="wgs_contig"
     misc_feature   193594. .194595
                    /note="wgs_contig"
ORIGIN

Alignment Scores:
Pred. No.:              77.7          Length:         222776
Score:                  74.00         Matches:        10
Percent Similarity:     100.00%       Conservative:   2
Best Local Similarity:  83.33%        Mismatches:     0
Query Match:            88.10%        Indels:         0
DB:                     14            Gaps:           0

US-09-858-332G-15 (1-12) x AC103448 (1-222776)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||:::||||||||:::
Db     183693 CATAACCACAACCACAACCACCACCACAACCACCAC 183658

RESULT 264
AC112472
LOCUS       AC112472               228998 bp    DNA     linear   HTG 20-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-77E16, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION   AC112472
VERSION     AC112472.5  GI:25138396
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 228998)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
```

Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 228998)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 228998)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Nov 20, 2002 this sequence version replaced gi:22856608.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

-------------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-------------- Project Information
Center project name: GSSC
Center clone name: CH230-77E16
-------------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 207321 bases at least Q40
Consensus quality: 210045 bases at least Q30
Consensus quality: 212155 bases at least Q20
Estimated insert size: 212371; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 227482: contig of 227482 bp in length
* 227483 227582: gap of unknown length
* 227583 228998: contig of 1416 bp in length.

FEATURES Location/Qualifiers
source 1. .228998
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-77E16"
misc_feature 1. .1539
/note="wgs_contig"
misc_feature 1864. .3160
/note="wgs_contig"
misc_feature 5840. .6898
/note="wgs_contig"
gap 227483. .227582
/estimated_length=unknown
ORIGIN

Alignment Scores:

Pred. No.:	79.8	Length:	228998
Score:	74.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	83.33%	Mismatches:	0
Query Match:	88.10%	Indels:	0
DB:	14	Gaps:	0

US-09-858-332G-15 (1-12) x AC112472 (1-228998)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||||||||:::|||||||||:::
Db     218166 CATAACCACAACCACAACCACCACCACAACCACCAC 218201
```

RESULT 265
AC095717/c
LOCUS AC095717 238494 bp DNA linear HTG 13-NOV-2002
DEFINITION Rattus norvegicus clone CH230-9K13, WORKING DRAFT SEQUENCE.
ACCESSION AC095717
VERSION AC095717.6 GI:24940991
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 238494)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,

```
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 238494)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 238494)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 13, 2002 this sequence version replaced gi:22773148.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GDDG
                Center clone name: CH230-9K13
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 221551 bases at least Q40
                Consensus quality: 223436 bases at least Q30
                Consensus quality: 224655 bases at least Q20
                Estimated insert size: 229610; sum-of-contigs estimation
                Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   238494: contig of 238494 bp in length.
FEATURES            Location/Qualifiers
     source         1. .238494
                    /organism="Rattus norvegicus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10116"
                    /clone="CH230-9K13"
     misc_feature   8872. .10696
                    /note="wgs_contig"
ORIGIN

Alignment Scores:
Pred. No.:              83              Length:         238494
Score:                  74.00           Matches:        10
Percent Similarity:     100.00%         Conservative:   2
Best Local Similarity:  83.33%          Mismatches:     0
Query Match:            88.10%          Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-15 (1-12) x AC095717 (1-238494)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||||:::|||||||||||||||:::
Db      69104 CACAACCACAACCACCACCACAACCACAACCACCAC 69069

RESULT 266
AF036170
LOCUS       AF036170               2400 bp    mRNA    linear   INV 20-DEC-1997
DEFINITION  Dictyostelium discoideum homeobox-containing protein Wariai mRNA,
            partial cds.
ACCESSION   AF036170
VERSION     AF036170.1  GI:2707267
KEYWORDS    .
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 2400)
  AUTHORS   Han,Z. and Firtel,R.A.
  TITLE     The homeobox-containing gene Wariai regulates anterior-posterior
            patterning and cell-type homeostasis in Dictyostelium
  JOURNAL   Development (1998) In press
REFERENCE   2  (bases 1 to 2400)
  AUTHORS   Han,Z. and Firtel,R.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-NOV-1997) Biology, UCSD, 9500 Gilman Drive, La Jolla,
            CA 92093, USA
FEATURES            Location/Qualifiers
     source         1. .2400
                    /organism="Dictyostelium discoideum"
                    /mol_type="mRNA"
                    /strain="KAx3"
                    /db_xref="taxon:44689"
     gene           1. .>2400
                    /gene="wariai"
     CDS            1. .>2400
                    /gene="wariai"
                    /note="stalk-cell specific; contains ankyrin repeats"
                    /codon_start=1
                    /product="homeobox-containing protein Wariai"
                    /protein_id="AAB92245.1"
                    /db_xref="GI:2707268"
                    /translation="MASIVMKSQHSLGYPNIGNINRSDYDSYEQQYNNPTGSKQYNNN
                    NNNNTNTNEINNGTNTNLNPNNMYGMYNNNNNNNNNNNNNNNNNNNNSNNNNNNNN
                    NNNINNNNNNNNNNNNNNNNNQHLSQSQQLSPTPYSSNSFSKLLSRSTNDLMLDQDD
```

PSKKKRKRTSPDQLKLLEKIFMAHQHPNLNLRSQLAVELHMTARSVQIWFQNRRAKAR
NMEFKPQLSHGGSDLIYNALGSQNGMNSMGGGGGNGGGNGGGGMNGINNILNGNHNRN
LNKYIPHGGNSINGNMGGGGGGGGGSHNHNHNHNHNHNHNHNHNQPLSNGDCGEKFS
VASAWNKILLHPNNIDFLIRYNPDDPNSIDVNARDSKGLSLLFTAAFLGYEYQVRRLI
ESGANPNIKDNQGDTPLIAASVLGNQPIVELLLEHRADPNLVNDEGVSPLFSACKGGH
LQIXSSLLDHDREVSVKTKINGETPLHIASLKGFEKICKLLIETEAKASVIDSNNRTP
LHHACIMGYFSIAKLLICNGADMNAIDIDGHTPLHTSSLMGQYLITRLLLENGADPNI
QDSEGYTPIHYAVRESRIETVKFLIKFNSKLNIKTKNGQNLIHLSVQFASLMMGQMIF
ESKGCEIAADDSDDQGYTPLYLAAKAGKTNFVKYLLSKGRSKKIRLEKLIQENQDKEI
IQMLESTVTKSSNNNNSNSNINNINNINNINNSQPNTNSDNNNNNNNNNFNENYSN
GNNEQSQPPGNKFEEDDEDDFYDRVYKKSYTNRIISNSFSYQQKLNSGNGISVNSGNL
ED"

ORIGIN

Alignment Scores:

Pred. No.:	1.15	Length:	2400
Score:	73.50	Matches:	12
Percent Similarity:	92.31%	Conservative:	0
Best Local Similarity:	92.31%	Mismatches:	0
Query Match:	87.50%	Indels:	1
DB:	2	Gaps:	1

US-09-858-332G-15 (1-12) x AF036170 (1-2400)

```
Qy          1 HisAsnHisAsnHisAsnHisAsn---HisAsnHisAsn 12
              ||||||||||||||||||||||||   ||||||||||||
Db        907 CATAATCATAATCATAATCATAATCATCACAATCACAAT 945
```

RESULT 267
DDDOKAGEN
LOCUS DDDOKAGEN 5284 bp DNA linear INV 18-APR-2005
DEFINITION D.discoideum dokA gene.
ACCESSION X96869
VERSION X96869.1 GI:1237201
KEYWORDS dokA gene; histidine kinase.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1
AUTHORS Schuster,S.C., Noegel,A.A., Oehme,F., Gerisch,G. and Simon,M.I.
TITLE The hybrid histidine kinase DokA is part of the osmotic response system of Dictyostelium
JOURNAL EMBO J. 15 (15), 3880-3889 (1996)
PUBMED 8670893
REFERENCE 2 (bases 1 to 5284)
AUTHORS Schuster,S.C.
TITLE Direct Submission
JOURNAL Submitted (25-MAR-1996) S.C. Schuster, Max-Planck-Institut fuer Biochemie Membranbiochemie, Am Klopferspitz 18a, Martinsried, 82152, FRG
FEATURES Location/Qualifiers
source 1. .5284
/organism="Dictyostelium discoideum"
/mol_type="genomic DNA"
/strain="AX2"
/sub_strain="214"
/db_xref="taxon:44689"
gene 44. .5056
/gene="dokA"
CDS 44. .5056
/gene="dokA"
/codon_start=1
/product="histidine kinase"
/protein_id="CAA65612.1"
/db_xref="GI:1237202"
/db_xref="GOA:Q23901"
/db_xref="InterPro:IPR000014"
/db_xref="InterPro:IPR001128"
/db_xref="InterPro:IPR001789"
/db_xref="InterPro:IPR003594"
/db_xref="InterPro:IPR003661"
/db_xref="InterPro:IPR004358"
/db_xref="InterPro:IPR005467"
/db_xref="UniProt/TrEMBL:Q23901"
/translation="MSSPHIELHSQRTLSPQPSSNNFELTGNKSCALSASLNGSIDDL
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNDGDKNDKKLLIITNAPTPP
LTPPQPTKTQQTLLELNQKFQEQQQQQQQQQQLQKQNKQQQSQKQQEEPPSSQQEEPP
LSQQQQEQEQEQEQEQEQEEQSKQKIEGKGGGGEEEECEGGGGGEGEGEEQFKEGDEE
ETRVLKLIQSTNNWIYWYRRSSLQINEASPKLIQTPSVDRTLCILDSLPCPIFQIGII
ENTENGNITLTGFANERLTTLLSCSSNLLVDSLKSIPSPHYNTVPSLESCNSIPINNL
APGLLPNTDPTKILDPYISLYNKKQLKERYSNNNNNTNTNNNNNNNNNNNNNNNNNNN
NNNNDTTTTTTTTTTTTTTTTTTTTTTTTTTNESDNNNNNNNNNNNNNNNNNNNNEE
IKQNTKEGEEILHESTNSTGTGNSISRPNSILQCILDSFHEQKTITEKVLLSNFLIRQ
QYKSRATIKPFPDGCICIFEYIENINLNYQQPPTSLNDRINSTNKLTNEMELLSLSPD
SNHQHVQKLHNHNHNHNHNHNHNNTTQRASSTDSPFIHSVSANSLSSMSSNSSVTSACA
SLSSNSSSNSNNNSNNASSSNCSSNANSNNTNSSSSNCCCCCNNNNNNNTSSTSSTIS
SASSTKSKSYLHQQHKNSKVYSTLLNSFENLSFLLKSTPIIIWRADHNGEMVFFKKSD
EIPIDEKKVVGNNFQDFLQWVPQTYRTNFQEMFQNSLKCGKIFEFLFWFQTPQNYVQL
FKIAGYPIFHIDGKDCIASKRYLIGWLGVTYNYLINDGAEISIKGSANLDSMYEKFKI
IYEMPNIERTKLTNNGISSGGDITNANNNGGSSESSNKILEKKKHLISEAEKECFLKC
KETYNILFKLSLLGVMFSTFKGIILDANDMLLQTIGYTRGDLENGKIDWMLLTPPEHF
EISARALQELKAKRWCQPIEKAYIHKSGKRVPVLITSAMIDGSTEQCITFVFDLSRYR
QAEMAAIEATRLKTQFITNISHELRTPCHGIVGMSQLLLDSQLTNTQRDNIDSIKRST
DSLISLINDILDFSKLEYGKVTLENESFELLPMIEEVLDSQATAANRKGIDLIFVMGR
DYPVPPVIFGDRNSLKKVLLNLVGNAVKFTETGFVLLEISTDYESGDQISLRFTVKDS
GIGIPENKIEQIFVPFGQIDGSFSRKYGGSGLGLSFCKELVALMGGYIRVESGDQEGG
KGTTFWFAIKVSISSPSYLPNSVPAANQFFYPEYRPSHYNNVIGNGDSPKNVLIIESN
QMVIMSIQSILLSMKCECISASKAITALDLLESSKSTENQIDIIVCSDKSTFVQQILD
YVTTEKVILYGVDPNSKYNENSKVYSYLVTPITHSKLISSILLSKNLKSKNSFLTTTN
NTNNTNNTNNIENKSSIDSPLSITSTSSSIITPILTSNNLDLNNNNNNNNNSILVSNN
GGVDGGNNVPSTLTTIQQSQPKKYILVAEDNDINIKVVVRQLEKLGYTAIVGINGLKA
LEIIGSFPICLILLDCQMPQMDGFTCSTILRQIEPTGQRIPIIAMTANDSKDRCFEVG
MDDYLSKPVRVDRLQKTLSDWIKTDENGNPTNSYNFYPLSYSLVYNNFIDTQLKKEKN
DD"

ORIGIN

Alignment Scores:

Pred. No.:	2.46	Length:	5284
Score:	73.50	Matches:	12
Percent Similarity:	92.31%	Conservative:	0
Best Local Similarity:	92.31%	Mismatches:	0
Query Match:	87.50%	Indels:	1
DB:	2	Gaps:	1

US-09-858-332G-15 (1-12) x DDDOKAGEN (1-5284)

```
Qy          1 HisAsnHisAsnHisAsn---HisAsnHisAsnHisAsn 12
              ||||||||||||||||||   ||||||||||||||||||
Db       1769 CATAATCATAATCATAATCATCATAATCATAATCATAAT 1807
```

RESULT 268
AC116957_0/c
WPCOMMENT
Sequence split into 4 fragments LOCUS AC116957 Accession AC116957

Fragment Name	Begin	End
AC116957_0	1	110000
AC116957_1	100001	210000
AC116957_2	200001	310000
AC116957_3	300001	405682

LOCUS AC116957 405682 bp DNA linear INV 12-MAR-2003
DEFINITION Dictyostelium discoideum chromosome 2 map 1685067-2090751 strain AX4, complete sequence.
ACCESSION AC116957 AC116962 AC117079
VERSION AC116957.2 GI:28829949
KEYWORDS HTG.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 405682)
AUTHORS Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P., Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K., Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and Noegel,A.A.
TITLE Sequence and analysis of chromosome 2 of Dictyostelium discoideum
JOURNAL Nature 418 (6893), 79-85 (2002)
PUBMED 12097910

```
  REMARK     The Dictyostelium Genome Sequencing Consortium
REFERENCE    2  (bases 1 to 405682)
  AUTHORS    Baumgart,C.
  TITLE      Direct Submission
  JOURNAL    Submitted (04-APR-2002) Genome Analysis, Institute of Molecular
             Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE    3  (bases 1 to 405682)
  AUTHORS    Baumgart,C.
  TITLE      Direct Submission
  JOURNAL    Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
             Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE    4  (bases 1 to 405682)
  AUTHORS    Baumgart,C.
  TITLE      Direct Submission
  JOURNAL    Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
             Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT      On or before Mar 4, 2003 this sequence version replaced
             gi:20066269, gi:19920061, gi:19919989.
             CDS predictions from GeneID do not necessarily reflect true genes.
             Further Information is available from IMB Jena, Department of
             Genome Analysis
             (http://genome.imb-jena.de/dictyostelium/)
             and the Univerity Colonge, Institute for Biochemistry I
             (http://www.uni-koeln.de/dictyostelium/project.shtml
             Funding
             Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES             Location/Qualifiers
     source          1. .405682
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /chromosome="2"
                     /map="1685067-2090751"
     CDS             complement(join(644. .776,858. .1605,1712. .1880))
                     /note="GeneID exon scores (in order of location ranges):
                     10.65, 75.03, 18.74 - GSCJ_ID dd_03093"
                     /codon_start=1
                     /product="similar to Arabidopsis thaliana (Mouse-ear
                     cress). Putative choline kinase"
                     /protein_id="AAO52440.1"
                     /db_xref="GI:28829950"
                     /translation="MMAKTKGSYEVFHYTVTKDKVNKGLCDIARYFVPEYRNSKDEDL
                     TIQKLNGGITNVLYLVEDKNIEQKYRYLPVVIRLYGYKSEEIIDRKNELIIQTEADQN
                     GLGAKFYGLFDNGCIYGFIKGEPLAYEDISKPTMQTCIAKEIAQWHSIEMPTRKNPSL
                     WPTIKKWAALAPDVYPVPEKNEYYQSINVKKMIEEGKMLEQRLAQLNSPIVFCHNDLL
                     SGNIIYDPSQNCASFIDFEYANYNFRGLELGNHFNEYAGFGPDYSLYPNKESQIHFLT
                     DYHRSLFKTEPTQDELEKLYIESNQFSLASHLYWGFWAIVQAMNSQIDFDYLEYGKAR
                     FDRYYETRDQFLNLN"
     CDS             join(3721. .5364,5457. .5528)
                     /note="GeneID exon scores (in order of location ranges):
                     133.16, 0.23 - GSCJ_ID dd_03097"
                     /codon_start=1
                     /product="similar to Dictyostelium discoideum (Slime
                     mold). Inositol 5-phosphatase 2"
                     /protein_id="AAO52441.1"
                     /db_xref="GI:28829951"
                     /translation="MGGNKQHHQACIQCAGVRFSKDEFGFLECDNCHTRVEGYFTQEV
                     EEFENGGGYKTGSYKRVWTQEVLDAMNQKKLRQKDNRYYILQNLLKIQVTSLIKDHGY
                     NEQLYQVTGELWFKLLDKFSQDFEGSSSTTSSSSSTTSTFGIINYPPPFITLSICYLA
                     CVYLREPVIISDFISLAQSGKIASFTNETTWSFSWSNSNSTFINFARISDDICKYLNF
                     LPSSNIPLLLLRFCTELKLPKSILNLSNQLYNLILIPFKLFRKKSPNSTLSKIYENIL
                     LPSNSDLLCVSIIMLILKSKFNLFDRQYNFNNSSGSNYNNNNNTLNDNNNNNNNNNEI
                     NNEYDINNIDKNYNKENSIERLEFLKKLINNNNFNSENSENSGISEERIKEEVEKMEL
                     ENWLYGVIYNLPKNEFLSKQDTWSLNIQDFKGIPKLKPSSIRRLITDQIHSKYDSLSD
                     KYQNLNNNNNNNNNNNNNNNNNNNIGNAINNIIKDDEGTLNEKIILKVKKSTTKVSTP
                     SKSKKSSLSSSSSSSSSSADPTILLPKVRNIRNAKSINRSGRTNLFPPSQKSVFSKGT
                     GKRIG"
     CDS             6218. .9748
                     /note="GeneID exon scores (in order of location ranges):
                     356.10 - GSCJ_ID dd_03095"
                     /codon_start=1
                     /product="similar to Dictyostelium discoideum (Slime
                     mold). Kinase responsive to stress 1-like kinase"
                     /protein_id="AAO52442.1"
                     /db_xref="GI:28829952"
                     /translation="MSNLKLSNNNNGNQKESSSPFQKVMKSPSTQNLLNSFSSNNSNN
                     NLSNSGSNEVKDTTTNSPSQLPPNYTPPPPPHQIRNSSSIEGGEFSLLNNENSDNNNN
                     NNNNNNNNNNNNNNNNNNNNNNNEQLARTESSVSIISSSSSGSNSGQPNLQRHSSNIS
                     TDDSNTTTETYSMSPNQTLNSNIDSSEQQHQDLSSSVNNNNNNNNNNNNNNNNNNNN
                     NNNNNNNNTHESRKLTRKIAQFISSPKLLQSSISQLPSTPTQQNVEIQTTNGGSSETS
                     PNGLISPRPSNDQPLKEKKKKKFLKTPEIFKHHHHHKESSLSSSTTTPSTTSSLTSSP
                     SSSSLAISSPNTTAATTTNKKSHKKTKSTFDINTEISVPYNVIHKMHVDFDLKWTGHN
                     DFILDEKLGDGAYGSVYKGTHKDLGFTLAIKVIEMKESESVSLQNEINILKNCKSPNI
                     VSYFGSLQTESHIWILLDFCALGSIRDIIESTEKTLNEAQISFVVKNTLKGLIYLHSQ
                     NIIHRDVKAANVLLSEGCDVKIADFGVSEKLNGALDQSKEMIGTPLWMAPEVILKKNY
                     DYKADIWSLGITIIEMADGLPPHIDLPPMRAMKMVPNWPPPTFAEPKKWSPLLNDFLA
                     RCLVKDPEKRASPIDLLCHPFLKKDRGPDVLSDLVNQLFKIKKKKIDDLKKQQKHQTS
                     QSSSSSSPQSPNATVNGGDIGSDGLSTSIISPIPSSPSDELDNCNNSLKLATSSKGML
                     SFKGNYTTCRDFQEEEEDSKHLNNNQDEQDDEQDDEDDDDENEDDEDVDPFSTTIFHG
                     KKKGSGNGNGGVTSDQDDEEEDEEEDDEEEEEEEEDDDEINEDEEISATGTMVVRKKK
                     NKSTKKSNKKKNKKNNLSTIGKSGSGNNLLHVAPNKPLPITPPFSISMTNEFKQLETK
                     LFTYIDSSNQKIVNDIKNEIKQLESSIINKININLQQQLSPILLALEEIKQNQHTTSQ
                     PKQMQSKLSATNLNEKKLSSSPPSSNSPLTNSVNSSLTTTTTTTTSPVLSRQSSFRSS
                     GSISSSNSSFRPNSAIMSAVNNSSTTTTTNSNSSSSNGGGSGVDISPTMTGRASPSIM
                     KRFTTSSSSSPLSSSSDGFAFNSNSNSSSSDLKRHVITPEELNNSNLVKNKVKMFEDD
                     NSSGSGSNSPSLSTNSSSTNSNNSVTNIKK"
     CDS             join(10002. .10168,10786. .12082)
                     /note="GeneID exon scores (in order of location ranges):
                     0.26, 60.31 - GSCJ_ID dd_00942"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO52443.1"
                     /db_xref="GI:28829953"
                     /translation="MESIFYSYLINNYDCGNNDEEFLNCLLILSINNLNITFNELLNY
                     CYNNNLKRKTLLNYFKNVKLIIEDNNGEIGFTPLTNRFYPLPYGVERVEVDCSSITIG
                     DHWIPDTCKSIYFHNFNQQKLNRNIISNSLTTIFFGNSFNQPLSDSNGLPWLPKKLKT
                     LSLGQSFQQTIQRNELPQSLTELFLDSEYQGIILNESIPKSVTTLHYEFFTIRKNHFD
                     TKSISKGVTSLKIGCYFNQTIKKDMIPINVTSLEIDNISKIEIEPNSLPLSIKTLRIG
                     SIELDSDNILPPNIKKLSIYGDYHSCFLPKSIESLKITSFFNLSKISFNNFRNLTSLK
                     IEKLSDSVLEEGVFPNTLLRIVLSCPNLKLSNGIFPNSLKKLKLKEFNHPIANGDLPS
                     SIEVLKLPNYDYPIKNGDLPSSLNVLKLQDHCISTIQRDALPPNLKLFYFHGRELRKL
                     PNLIWPTSFEFAYISEASFNFINSIDLDFFTNHIRLK"
     CDS             join(12253. .12311,12539. .13007)
                     /note="GeneID exon scores (in order of location ranges):
                     -5.63, 15.15 - GSCJ_ID dd_00941"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO52444.1"
                     /db_xref="GI:28829954"
                     /translation="MDNIIYSYLINNYNNKNNEEQEEEEKKHLNDLKGLNDQQFQTLN
                     NHIQNLKNTFTTNNSNDEVIYNHSITDNNNNSNDTINNRNKRIAFDCSILIIIKNNLN
                     HHITIKNNDNSNNLNQLILKPQPDLNNKIRTNQQIDQTIDNNYKLFKKVWGNIFIRDQ
                     ILFHLRIYNNHNHSI"
     CDS             join(13144. .14052,15655. .15768)
                     /note="GeneID exon scores (in order of location ranges):
                     36.56, 7.78 - GSCJ_ID dd_00940"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO52445.1"
                     /db_xref="GI:28829955"
                     /translation="MRCSDIILVDHWIPDTCKSIIFSHFNQQLTRNIIGNSLTSIDFG
                     KDFNQSLSDKNGIPWLPRTLKSLTFGKSFQHSIHRDEIPPSLTSLILDPKYKGVIEYG
                     SIPKSVTTLHYYFDSKGGAKLDSTSIPGGTTSLEFDSFFDQTIEKGMISHNVTSLKFS
                     FCNSIIKSNSLPISIKTLHLGSNQNFPRDVFPQAITNLSVYTFSDIKILPKTIQKLKI
                     MDFSPPSQLDNHSFYFLKNLTSLNFDAKQVDLTNVKFPNIISRLTLAVSDYLEFNDCL
                     LQNNFKKLKLFNFNQPLNVGDLPSSIEGCPAVRAIRANGLIQALLSDHGLDVNNLNIT
                     FNQLQD"
     CDS             join(15826. .16011,16072. .17106)
                     /note="GeneID exon scores (in order of location ranges):
                     1.92, 31.02 - GSCJ_ID dd_00422"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO52446.1"
                     /db_xref="GI:28829956"
                     /translation="MTQLTKRNKRIAFDCSILIIIKNNLNHHITIKNNLNHTTNNSNN
```

```
                     KCNNLNQVVILKLIDINKFGEIYLSGIKFYFILEYIIIRIIYSIDIGGHNISTYKYRN
                     YFNNSNISNSDENNNHQQNISNCCPMPSTAEKIEMNCSDIILVDHWIPDTCKSIKFTY
                     FDQQLTRNIIGNSLTSINFSNEFNQSLSDTKGEPWLPRTLKSLTFGKSFQHSIHRDEM
                     PPSLTTLIFIPVYKGVIQYGSIPKSVTTLHYSLFFKEHDIYKFDLTSVPGGTTSLEFD
                     FYYNGTIEKGMIPHYVTSLKFRMGDYKLEPNSLPISIKTLGLGRNRTIFTQKVFPPTI
                     TNLSIVTFSDIGILPKTIEKLKIIGSNNFTQLDSHSFHIFENLKSLELNAIGVDLTGV
                     EFSPTIIKLTLTVL"
     CDS             join(18782. .18911,19094. .20841)
                     /note="GeneID exon scores (in order of location ranges):
                     0.11, 70.39 - GSCJ_ID dd_00420"
                     /pseudo
                     /codon_start=1
     CDS             join(21335. .21528,21616. .21928)
                     /note="GeneID exon scores (in order of location ranges):
                     -5.52, 10.92 - GSCJ_ID dd_00435"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO52447.1"
                     /db_xref="GI:28829957"
                     /translation="MDSIIYFYLINNYNNKNNEELLNCLLILSINNLNITFNQVQDYC
                     NSINNYKNNYQNLQKKKENKLQEEEEETHLNDLKGLNDQQFQTLNNHIQNLKNTFITN
                     NSNDETIYTHSITDNNNNDTIDNRNKRIAFDCSILIIIKNNLNHHITINNNKNINNNN
                     NLYQLILK"
     CDS             complement(22154. .22339)
                     /note="GeneID exon scores (in order of location ranges):
                     0.40 - GSCJ_ID dd_00433"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO52448.1"
                     /db_xref="GI:28829958"

Alignment Scores:
Pred. No.:                45.7          Length:          110000
Score:                    73.50         Matches:         12
Percent Similarity:       92.31%        Conservative:    0
Best Local Similarity:    92.31%        Mismatches:      0
Query Match:              87.50%        Indels:          1
DB:                       2             Gaps:            1

US-09-858-332G-15 (1-12) x AC116957_0 (1-110000)

Qy          1 HisAsnHisAsnHisAsn---HisAsnHisAsnHisAsn 12
              |||||||||||||||||   |||||||||||||||||||
Db      93164 CATAATCATAATCATAATCATCATAATCATAATCATAAT 93126

RESULT 269
AC094631/c
LOCUS       AC094631               254816 bp    DNA     linear   HTG 09-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-5D14, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION   AC094631
VERSION     AC094631.6  GI:30466395
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 254816)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 254816)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 254816)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 9, 2003 this sequence version replaced gi:24818467.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            ---------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            ---------- Project Information
                Center project name: GBBI
                Center clone name: CH230-5D14
            ---------- Summary Statistics
                Assembly program: Atlas;
                Consensus quality: 243727 bases at least Q40
```

```
            Consensus quality: 245833 bases at least Q30
            Consensus quality: 247014 bases at least Q20
            Estimated insert size: 254075; sum-of-contigs estimation
            Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
          --------------
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
          * NOTE: This sequence may represent more than one clone.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 2 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1   243433: contig of 243433 bp in length
          *   243434   243533: gap of unknown length
          *   243534   254816: contig of 11283 bp in length.
FEATURES             Location/Qualifiers
     source          1. .254816
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-5D14"
     misc_feature    1. .1743
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    2955. .3828
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH335197"
     misc_feature    3138. .6370
                     /note="wgs_contig"
     misc_feature    7408. .8246
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH335199"
     misc_feature    242368. .243433
                     /note="wgs_end_extension
                     clone_end:Sp6"
     gap             243434. .243533
                     /estimated_length=unknown
     misc_feature    247771. .249172
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    253186. .254816
                     /note="wgs_end_extension
                     clone_end:Sp6"
ORIGIN

Alignment Scores:
Pred. No.:              103          Length:        254816
Score:                  73.50        Matches:       12
Percent Similarity:     92.31%       Conservative:  0
Best Local Similarity:  92.31%       Mismatches:    0
Query Match:            87.50%       Indels:        1
DB:                     14           Gaps:          1

US-09-858-332G-15 (1-12) x AC094631 (1-254816)

Qy          1 HisAsnHisAsnHis---AsnHisAsnHisAsnHisAsn 12
              ||||||||||||||   ||||||||||||||||||||
Db     241771 CACAACCACAACCACAACAACCACAACCACAACCACAAC 241733

RESULT 270
AY084699/c
LOCUS       AY084699                 862 bp    mRNA    linear   PLN 27-MAR-2003
DEFINITION  Arabidopsis thaliana clone 11539 mRNA, complete sequence.
ACCESSION   AY084699
VERSION     AY084699.1  GI:21403409
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 862)
  AUTHORS   Haas,B.J., Volfovsky,N., Town,C.D., Troukhan,M., Alexandrov,N.,
            Feldmann,K.A., Flavell,R.B., White,O. and Salzberg,S.L.
  TITLE     Full-length messenger RNA sequences greatly improve genome
            annotation
  JOURNAL   Genome Biol. 3 (6), RESEARCH0029 (2002)
   PUBMED   12093376
REFERENCE   2  (bases 1 to 862)
  AUTHORS   Brover,V., Troukhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and
            Feldmann,K.
  TITLE     Full-Length cDNA from Arabidopsis thaliana
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 862)
  AUTHORS   Brover,V., Troukhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and
            Feldmann,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road,
            Malibu, CA 90265, USA
COMMENT     This clone sequence is one of 5,000 Ceres full-length cDNAs made
            available to TIGR and Genbank. The following quality assessment of
            this set was done by comparison with known proteins: two percent of
            the clones are estimated to be 5'-truncated; less than one percent
            are 3'-truncated; approximately two percent represent alternative
            splice variants, including unspliced introns and spliced exons; one
            percent may contain premature stop codons; five percent may have
            frame shifts in a coding region. A sequence is considered to be
            5'-truncated if it lacks the translation initiation start (ATG). A
            sequence is considered to be 3'-truncated if it lacks the
            C-terminal end of the encoded protein. Please note that these cDNA
            sequences are derived from the Ws or LAer ecotypes and therefore
            may contain polymorphisms when compared to sequences from Col-0.
            Genset carried out the library production and sequencing of the
            full-length clones. Ceres, Inc. carried out the clustering of the
            5' sequences, selection of clones, and sequence assembly.
FEATURES             Location/Qualifiers
     source          1. .862
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="11539"
     CDS             60. .689
                     /codon_start=1
                     /product="unknown"
                     /protein_id="AAM61260.1"
                     /db_xref="GI:21536928"
                     /translation="MKREGKQHGMVRTYRILPPSLNPRPESKLVNPLTSRPTAGLFTK
                     VTSKPTSHSKFTGKCGQARCLECHMHPVTKSKAKTKGSSKVRTNDVTYKMLTWQVAAG
                     GPRPGMKLSGFSATGILDLMSDDYGYDHDYDYDEEDEEEENRGSVVEEVVKIQSSDDG
                     GETEEDGSHDDDDTDDDGRMSFCDVGMMMMMDHVEEYDEGWYLVEEMMT"
ORIGIN

Alignment Scores:
Pred. No.:              0.497        Length:        862
Score:                  73.00        Matches:       10
Percent Similarity:     100.00%      Conservative:  1
Best Local Similarity:  90.91%       Mismatches:    0
Query Match:            86.90%       Indels:        0
DB:                     15           Gaps:          0

US-09-858-332G-15 (1-12) x AY084699 (1-862)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              ||||||||||||||:::|||||||||||||||
Db        462 CATAATCATAATCATGATCATAACCATAATCAT 430

RESULT 271
AC013163
```

```
LOCUS       AC013163               14923 bp    DNA     linear   HTG 03-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC013163
VERSION     AC013163.1  GI:6223168
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 14923)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10213807 by the submitter.
            For further information on this sequence you may e-mail to
            fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .14923
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
ORIGIN

Alignment Scores:
Pred. No.:              7.75          Length:        14923
Score:                  73.00         Matches:       10
Percent Similarity:     100.00%       Conservative:  1
Best Local Similarity:  90.91%        Mismatches:    0
Query Match:            86.90%        Indels:        0
DB:                     14            Gaps:          0

US-09-858-332G-15 (1-12) x AC013163 (1-14923)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              |||||||||||||||||||||||||||:::|||
Db      12875 CACAATCACAATCACAATCACAATCACGATCAC 12907

RESULT 272
AC156596/c
LOCUS       AC156596               39074 bp    DNA     linear   HTG 31-JAN-2005
DEFINITION  Laccaria bicolor clone JGIAHZW-15X, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION   AC156596
VERSION     AC156596.1  GI:58338252
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Laccaria bicolor
  ORGANISM  Laccaria bicolor
            Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
            Agaricales; Tricholomataceae; Laccaria.
REFERENCE   1  (bases 1 to 39074)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Unpublished
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 39074)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2005) Production Genomics Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA
            94598-1698, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 3640135
            Center clone name: JGI-AHZW_15X
            -------
            Summary Statistics
            Consensus quality: 38787 bases at least Q40
            Consensus quality: 38932 bases at least Q30
            Consensus quality: 38960 bases at least Q20
            Estimated insert size: 40000; agarose-fp estimation
            Estimated insert size: 38974; sum-of-contigs estimation
            Quality coverage:  7.26 in Q20 bases; agarose-fp estimation
            Quality coverage:  7.45 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    11646: contig of 11646 bp in length
            *    11647    11746: gap of unknown length
            *    11747    39074: contig of 27328 bp in length.
FEATURES             Location/Qualifiers
     source          1. .39074
                     /organism="Laccaria bicolor"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:29883"
                     /clone="JGIAHZW-15X"
                     /clone_lib="JGI Fosmid library AHZW"
     gap             11647. .11746
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              19.6          Length:        39074
Score:                  73.00         Matches:       10
Percent Similarity:     100.00%       Conservative:  1
Best Local Similarity:  90.91%        Mismatches:    0
Query Match:            86.90%        Indels:        0
DB:                     14            Gaps:          0

US-09-858-332G-15 (1-12) x AC156596 (1-39074)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              |||||||||||||||||||||||||||:::|||
Db      12612 CACAACCACAACCATAACCACAACCACCACCAC 12580

RESULT 273
DMBN4L24/c
LOCUS       DMBN4L24               71813 bp    DNA     linear   INV 10-OCT-1999
DEFINITION  Drosophila melanogaster clone BACN4L24.
ACCESSION   AL121805
VERSION     AL121805.1  GI:6018875
KEYWORDS    .
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 71813)
  AUTHORS   Mottier, Cadieu, Dreano, Lelaure and Galibert,F.
  TITLE     Sequencing the distal X chromosome of Drosophila melanogaster
  JOURNAL   Unpublished
  REMARK    UPR41 CNRS Recombinaisons Genetiques, Universite de Rennes 1 2 av
            du Pr Leon Bernard,Faculte de Medecine, 35043 RENNES Cedex, France.
REFERENCE   2  (bases 1 to 71813)
  AUTHORS   Benos,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-OCT-1999) European Drosophila Genome Sequencing
            Consortium
COMMENT     Sequence submitted by Takis Benos, EMBL Outstation - The EBI,
            Hinxton Hall, Hinxton, Cambridge, CB10 1SD, U.K.
            E-mail: benos@ebi.ac.uk on behalf of the European Drosophila Genome
            Sequencing Consortium. For further information see the European
            Drosophila Genome Sequencing Consortium's web site:
```

```
          http://edgp.ebi.ac.uk/.
          This is the finished sequence of clone BACN4L24.  The CDSs
          presented here are only preliminary predictions based on automatic
          analysis. Analysis of the sequence is in progress.
          Clone=BACN4L24; Contig ID=0; Length=71813; Status=Finished.
FEATURES             Location/Qualifiers
     source          1. .71813
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN4L24"
ORIGIN

Alignment Scores:
Pred. No.:                35.2          Length:          71813
Score:                    73.00         Matches:         10
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    90.91%        Mismatches:      0
Query Match:              86.90%        Indels:          0
DB:                       2             Gaps:            0

US-09-858-332G-15 (1-12) x DMBN4L24 (1-71813)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              ||||||||||||||||||||||||||:::|||
Db      18295 CACAATCACAATCACAATCACAATCACGATCAC 18263

RESULT 274
AB009052/c
LOCUS       AB009052               85992 bp    DNA     linear   PLN 14-FEB-2004
DEFINITION  Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone:MNF13.
ACCESSION   AB009052 BA000015
VERSION     AB009052.1  GI:2656028
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Sato,S., Kaneko,T., Kotani,H., Nakamura,Y., Asamizu,E., Miyajima,N.
            and Tabata,S.
  TITLE     Structural analysis of Arabidopsis thaliana chromosome 5. IV.
            Sequence features of the regions of 1,456,315 bp covered by
            nineteen physically assigned P1 and TAC clones
  JOURNAL   DNA Res. 5 (1), 41-54 (1998)
    PUBMED  9628582
REFERENCE   2  (bases 1 to 85992)
  AUTHORS   Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-NOV-1997) Yasukazu Nakamura, Kazusa DNA Research
            Institute, Department of Plant Gene Research; 1532-3, Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
            Tel:81-438-52-3935, Fax:81-438-52-3934)
COMMENT     Address for correspondence: kaos@kazusa.or.jp
            For the latest information on annotation of this clone, please see
            http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MNF13
            Genes with similarity to proteins in the databases are described in
            'product' or 'note' qualifiers.  Genes that have no significant
            protein similarity are described as 'unknown protein'.
            The software programs used to predict genes include: Grail
            (Informatics Group, Oak Ridge National Laboratory,
            http://compbio.ornl.gov/Grail-1.3/),
            GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
            NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
            Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
            SplicePredictor (Volker Brendel, Stanford University,
            http://gremlin1.zool.iastate.edu/cgi-bin/sp.cgi).
            Genes encoding tRNAs are predicted by tRNAscan-SE
            (Sean Eddy, Washington University School of Medicine, St. Louis,
            http://genome.wustl.edu/eddy/tRNAscan-SE/).
            This sequence may not be the entire insert of this clone. It may be
            shorter because we remove overlaps between neighboring submissions.
            The 5' clone is K21I16 and the 3' clone is K1B16.
FEATURES             Location/Qualifiers
     source          1. .85992
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /chromosome="5"
                     /clone="MNF13"
                     /clone_lib="Mitsui P1"
                     /ecotype="Columbia"
     CDS             join(244. .368,668. .708,812. .975,1066. .1165,1255. .1424,
                     1513. .2184)
                     /note="unnamed protein product; contains similarity to RNA
                     binding protein
                     gene_id:MNF13.1"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAB08520.1"
                     /db_xref="GI:9758076"
                     /translation="MDPYGTETVIEDEIHDPKPSEDIEDDDDKSQPHSGGGVDSAGKI
                     FVGGLARETTSAEFLKHFGKYGEITDSVIMKDRKTGQPRGFGFVTYADSSVVDKVIQD
                     NHIIIGKQVEIKRTIPRGSMSSNDFKTKKIFVGGIPSSVDDDEFKEFFMQFGELKEHQ
                     IMRDHSTGRSRGFGFVTYESEDMVDHLLAKGNRIELSGTQVEIKKAEPKKPNSVTTPS
                     KRFGDSRSNFGGGYGDGYGGGHGGGYGGPGGPYKSGGGYGGGRSGGYGGYGGEFGGYG
                     GGGYGGGVGPYRGEPALGYSGRYGGGGGGYNRGGYSMGGGGGYGGGPGDMYGGSYGEP
                     GGGYGGPSGSYGGGYGSSGIGGYGGGMGGAGGGGYRGGGGYDMGGVGGGGAGGYGAGG
                     GGNGGGSFYGGGGGRGGYGGGGSGRYHPYGR"
     CDS             join(4110. .4271,4372. .4445,4682. .4987,5065. .5148,
                     5256. .5559)
                     /note="gene_id:MNF13.2"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="sucrose cleavage protein-like"
                     /protein_id="BAB08521.1"
                     /db_xref="GI:9758077"
                     /translation="MYSTNIANSITSYARHVFVLYKTPEAWLSHVEEEGLPQRFATLL
                     KDRKSDLLVQTKLNVCEGGGSDGDVLIFPDMIRYKGVKDTDVEGFFEDVLVNGKPWSS
                     GIQEEISGTFVFVCTHASRDKRCGVCGPVILERFKQEIGSRGLSDQITLKRCSHVGQH
                     KYAGNLIIFCPDSAGKITGNWYGYVTPDDVPELLDQHIAKGEIIQRIWRGQMGLPGGE
                     AEKVHEQKVIPNGHGVVKEESKGFTGGCCQGSNGVSCCQDETPKPEPIKKEVKKCTIW
                     FQPLDKEELYIGAAVVGAIATIAMAFTFFKRSG"
     CDS             complement(join(6276. .6315,6416. .6509,6717. .8472,
                     8550. .8583,8670. .8710))
                     /note="unnamed protein product; gene_id:MNF13.3
                     unknown protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAB08522.1"
                     /db_xref="GI:9758078"
                     /translation="MSDSTKDHTPGKTEFVPESLLREDDIFYETIKSRVEEAFGFCKN
                     QQVVCNDVQQKELKFCDLELVRKLDSLTNKGLYLIAMILTGGSTSFDKTRLKMKEIIR
                     DSVRRDFGKNKEGIGKEEIINQLHQVLSDPANFREDCRMNLGRTPTLHSHRDAAMKVL
                     NELDSLSTQTLRAMKRKLKGSRMIPQLKNSRRGQRRSDLINQVRQASEKMLSELSAGD
                     KLQEQLAKALSVVDLSLKLSPGYKTAAATDFFRFSPETKNLQNEIVKAVWLLRKVRFR
                     ELKRLHLCLDPEAEVSNDSLRSAVRKTLIEYLFECSDLDTIPKSLMEALSLVNSRTGN
                     VEHKVCPREAIEEETECILNVSAQVKQIFCQCIPNYELDQDFGDAYMEDLEDSDDNDD
                     DDDDDDDDDENCNLFDNEKLGNKEESRCRNIKLEVKDTQDDAMESSDSDHEESGAECL
                     VLDPTDSTHETNQHDFSSSVNRVVVRDLPESITRVYPRSLYVTPTSNKSIVISDRHDI
                     AMSKTRVKVERDIEMEVDNQFSSRSLFSVENIKSDDHGEQKPRRRNKNQYLAIQEISD
                     ETSLVAHNLIGRLLEKFADRQRLNLETDERSYLGGESRLQEKVEVSEKKQASSQAKSD
                     EAIILSVVKEQITSLEESVLMRLKELMEVS"
     CDS             join(9312. .9512,9598. .9659,9739. .9790,9876. .9990,
                     10076. .10170,10260. .10366,10460. .10688,10790. .10861)
                     /note="unnamed protein product; contains similarity to
                     unknown protein
                     dbj|BAA23705.1
                     gene_id:MNF13.4"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAB08523.1"
                     /db_xref="GI:9758079"
                     /translation="MTTEENKTSRNRKRKRQRNPKPSKEEPIETTPKNQNEKKNQRDT
                     KNQQHGGSSAPSKRPKPSNFLDALRERLSGGQFRMLNEKLYTCSGKEALDYFKEDPQM
```

FDMYHTGYQQQMSNWPELPVNSIINWLLSNSSSLVVADFGCGDARIAKSVKNKVFSFD
LVSKNPSVIACDMSNTSLESSSVDVAVFCLSLMGTNYSSYIKEAHRVLRPSGMLLIAE
VKSRFDPNNGGADPKDFVKAVCDLGFTSVLKVQSFLLLFFISTIRKKLIWFFEQDFSN
KMFILFHFKKKEQMNSNQKIIKWPELKACLYKRR"

CDS join(12391. .12521,12945. .13077,13176. .13371,13448. .13572,
13660. .13818,13914. .14231)
/note="gene_id:MNF13.5"
/codon_start=1
/evidence=not_experimental
/product="protein kinase ATN1"
/protein_id="BAB08524.1"
/db_xref="GI:9758080"
/translation="MGSVTGFYSNEVFELDPKWVVDPQHLFVGPKIGEGAHAKIYEGK
YKNKTVAIKIVKRGESPEEIAKRESRFAREVSMLSRVQHKNLVKFIGACKEPIMVIVT
ELLLGGTLRKYLVSLRPGSLDIRVAVGYALDIARAMECLHSHGVIHRDLKPESLILTA
DYKTVKLADFGLAREESLTEMMTAETGTYRWMAPELYSTVTLRHGEKKHYNHKVDAYS
FAIVLWELIHNKLPFEGMSNLQAAYAAAFKNVRPSADDLPKDLAMIVTSCWKEDPNDR
PNFTEIIQMLLRCLSTISSTELVPPAIKRVFSSENTVLPPESPGTCSLMTVRDKDQIP
TDANSAQNEVRGSFFFFCC"

tRNA complement(14396. .14470)
/product="tRNA-Asp"
/note="codon recognized: GAC; gene_id:MNF13.6"
/evidence=not_experimental

CDS join(16046. .16163,16411. .16504,16638. .16717,16805. .16914,
16988. .17073,17258. .17311,17393. .17546,17622. .17861)
/note="unnamed protein product; dbj|BAA95721.1
gene_id:MNF13.7
strong similarity to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB08525.1"
/db_xref="GI:9758081"
/translation="MSSPDIVGILENTKELDRLRKDQEEVLVEINKMHKKLQASPEIV
EKPGDISLAKLKNLYIQAKELSENEVTVSNILLTQLDLLLPYGPTGQQRRKLGEGNDQ
KRKRMKVDSDVIRLSPSMRNQIEAYASLKGEQVAARVTAESADKDEWFVVKVIHFDRE
TKEVEVLDEEPGDDEEGSGQRTYKLPMLCILPFPKRNDPSNTQEFPPGKHVLAVYPGT
TALYKATVVSTPRKLLNMDVNSSNIRGSLTSKFVNLLCLLYTENGVFSSISYNKTICL
IRYLLEFDDDEEDGALPQRTVPFHKVVALPEGHRQ"

CDS join(18298. .18378,18988. .19114,19284. .19529,19694. .19889,
19989. .20245,20341. .20468,20578. .20712,20908. .20958,
21041. .21280)
/note="unnamed protein product; contains similarity to
serine protease
gene_id:MNF13.8"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB08526.1"
/db_xref="GI:9758082"
/translation="MIQLNEPLTHVRTKPILISLEFESPGRKTIDSVATESVLNSVVK
INTFSSKPNICYPWQNKPQKKSKGSGFVIPGKMIITNAHVVANHILVLVIKRGSPKKY
KAEVKAIGRECDLAILVIESKEFWEDMNPLELGDMPFLQESVNVIGYPTGGENISVTK
GVVSRIESMDYAHGAINLPAIQTDAAMNPGNSGGPVCIGNKVVGVAFQTLGHSNNIGC
LIPAPVVKHFITGVEKTGQYVGFCSLNLSYQHMDAQTRSHFKMNSEMTGILIYNINQH
SDALNILKKYDVILSIDGVAIENDGTVIIPNRERTRLDDLVSLKQLGETILLKILREG
KMHEFNITLRPVQRLVPAGQIDNNPSYYIFAGFVFVPLRKQHFKGSNGEQIVVISEVL
ADVINVEYYMYKHLKVNSVNKVKVENLKHLCELIEKCCTKDLRLELGDGRVIILDYQF
AKSSTSLILERHRVPWAMSKDLMTDQSTTCSSTRLY"

CDS join(21559. .21782,22349. .22446,22555. .22722,22802. .22844,
22931. .23012)
/note="unnamed protein product; emb|CAB87783.1
gene_id:MNF13.9
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB08527.1"
/db_xref="GI:9758083"
/translation="MMKEKAKEGKSLFGSPTFVDLGNGRLRCVETGHEVLAGDVESYA
RNKRCRLGLINHALSLGKPPLNMFSQCPLSRSKLVCKLTGDTVNKNEEHIWKHVNGKR
FLHKLEQVERGAGSSGRTEETQVNNPRHNKEDSGSEESDFWMPKSSSDSELDEETDEE
NCKGSHCDDKESEELSERTKRMSLEIGPSSFASRKKKSKNDELS"

CDS complement(join(23272. .23351,23476. .23563,23740. .23838,
23922. .24026,24126. .24195,24277. .24373,24499. .24640,
25069. .25212))
/note="gb|AAC32067.1
gene_id:MNF13.10"
/codon_start=1
/evidence=not_experimental
/product="20S proteasome beta subunit PBB2"
/protein_id="BAB08528.1"
/db_xref="GI:9758084"
/translation="MSQSSVDIPPKGGFSFDLCKRNDMLTQKGLKAPSFLKTGTTIVG
LIFKDGVILGADTRATEGPIVADKNCEKIHYMAPNIYCCGAGTAADTEAVTDMVSSQL
RLHRYQTGRDSRVVTALTLLKKHLFSYQGHVSAALVLGGVDITGPHLHTIYPHGSTDT

Alignment Scores:

Pred. No.:	41.9	Length:	85992
Score:	73.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	90.91%	Mismatches:	0
Query Match:	86.90%	Indels:	0
DB:	15	Gaps:	0

US-09-858-332G-15 (1-12) x AB009052 (1-85992)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              ||||||||||||||:::|||||||||||||||
Db      65933 CATAATCATAATCATGATCATAACCATAATCAT 65901
```

RESULT 275
AC116924/c
LOCUS AC116924 88549 bp DNA linear INV 21-FEB-2004
DEFINITION Dictyostelium discoideum chromosome 2 map 6357117-6445670 strain AX4, complete sequence.
ACCESSION AC116924
VERSION AC116924.3 GI:42733690
KEYWORDS HTG.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 88549)
AUTHORS Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P., Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K., Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and Noegel,A.A.
TITLE Sequence and analysis of chromosome 2 of Dictyostelium discoideum
JOURNAL Nature 418 (6893), 79-85 (2002)
PUBMED 12097910
REMARK The Dictyostelium Genome Sequencing Consortium
REFERENCE 2 (bases 1 to 88549)
AUTHORS Baumgart,C.
TITLE Direct Submission
JOURNAL Submitted (03-APR-2002) Genome Analysis, Institute of Molecular Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE 3 (bases 1 to 88549)
AUTHORS Baumgart,C.
TITLE Direct Submission
JOURNAL Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE 4 (bases 1 to 88549)
AUTHORS Baumgart,C.
TITLE Direct Submission
JOURNAL Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE 5 (bases 1 to 88549)
AUTHORS Baumgart,C.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2004) Genome Analysis, Institute of Molecular Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT On Feb 21, 2004 this sequence version replaced gi:28829074.
CDS predictions from GeneID do not necessarily reflect true genes.
Further Information is available from IMB Jena, Department of Genome Analysis
(http://genome.imb-jena.de/dictyostelium/)
and the Univerity Colonge, Institute for Biochemistry I
(http://www.uni-koeln.de/dictyostelium/project.shtml
Funding

```
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES            Location/Qualifiers
     source         1. .88549
                    /organism="Dictyostelium discoideum"
                    /mol_type="genomic DNA"
                    /strain="AX4"
                    /db_xref="taxon:44689"
                    /chromosome="2"
                    /map="6357117-6445670"
     CDS            1864. .2241
                    /note="GeneID exon scores (in order of location ranges):
                    27.07 - GSCJ_ID dd_01259"
                    /codon_start=1
                    /product="hypothetical protein"
                    /protein_id="AAO51640.2"
                    /db_xref="GI:42733691"
                    /translation="MDQFKVPPPTYAPVAGQTIYGAPSYNNYYMRTPQVSTAPMIYPT
                    PMMAPPIMTPPIMTQPIMTPPMMYPPIIPSQPFMGPSMVSPIMSPPMIPSQPFMGPSM
                    VSPGYGVTPNLNVPFSTNVKYYD"
     CDS            2839. .4479
                    /note="GeneID exon scores (in order of location ranges):
                    88.56 - GSCJ_ID dd_01260"
                    /codon_start=1
                    /product="hypothetical protein"
                    /protein_id="AAO51641.2"
                    /db_xref="GI:42733692"
                    /translation="MDCKIHNEKNIMWCLNCTTLCCVDCIEFDHFQHNLKSFKNIDKF
                    HCPTHDEPIIKYCILCNCVICNLCEKTTHFHNNNNNQINITQFPWIPKKKSKSIELEL
                    NEIDEYNNNYNNNINNNNNNNNKNNVNNVNNVNNENKEINENHNNKNQRSNSPPPLPK
                    IKLLNIENITLIKHGSYRNLNKIEEENENDEKEILDRKQYNKENKEFLKKQSEKGNSI
                    AQCLLGSSYYYGMNGYSVNFEKSLKYLTKSAEQNNAKAYFELFNYYKQQKKDLIKSKY
                    YLLKSAEFIYEDSIILLGDYYYNGFKTIDNHDDHDDGSGENEELNLICQDYFKAFNLF
                    KIGASLNVDSNSLTKLGICFYYGRGVTIDYNEAYRLFYQSSICDHHQEHNPEITFYYL
                    GLCLFYGKGVLKNQCKGFEYFMKSASLNHYSPALEAVGRCFLNGEGISQNFMQAKLYF
                    TTAKSQGSNQDRNIQDVNKIIELNKSIKIKSPNILIEIEKIINENKSFNLEQSCLNYI
                    NHQKNNYKQEIDNTNIKNIKCKQDSLIDKTFNKVNYLN"
     CDS            complement(join(4673. .5239,5354. .6118,6205. .6621,
                    6720. .7982))
                    /note="GeneID exon scores (in order of location ranges):
                    34.62, 63.42, 31.78, 100.55 - GSCJ_ID dd_01261"
                    /codon_start=1
                    /product="hypothetical protein"
                    /protein_id="AAO51642.2"
                    /db_xref="GI:42733693"
                    /translation="MEIDKNVEQSSNKVEYNQPKSPITSILVNNNSSNNNNNSNNKNN
                    NNEEEFDISSPLTLSPIANCGASGSNIDSSIDNTPHLLSQISTPPISSTSSITLLNDK
                    KKGIYSTFLDLDYYNNTYLKNKQKILQPKLNQQKQQQQQQQQQPKTLTFSLDSQQQQQ
                    QQPQLQSPQLIDQQQQQQQQQQQQQQQQQQQQQQQIPYHQRKHHISIRDGDDNDNEVDS
                    FNILSQNQLIHQTFNNPQNHHYFSSTSPSLLYHHQQQTQLEQKPIYSSSSSSSILSSL
                    SSSTISFYNSITSLFLTFPQTISTDPIIIREYERLKMLQEQQVGPRHTEILLRKEDGV
                    GGRKALKEKFVEIYESFFNGEDPSENQPIFWEQLFLIKVNIPPLERCIILTSEENLLA
                    LKQNINKIFTQSCFWLKDSNSLRLTHIIETLSILLKSIPRKKFNNFGFDILNILCGIE
                    NSDQVFTDMIKDLEQLLKNSSIDKKVKIGVINLFTIIATATEYINQNTLLQYFMSVNI
                    FDTLVTFLGEDDIPKQTQLNILTLLCYFSNFQKYEITNPYLKDLSNLSDPIKILKILN
                    LITQNLKDHNEFFEQQYRDPNLGMVSRFSGYLSSWVYSPKVVQPLSFYETGSSLMILY
                    ELFYQNDHFVNMLVKGGLPFANGGENGSTCPDIIKQFFSFCAYLASDAQKDLKTIYSR
                    MALLILLCICEKRELEDFFHDIKSTCYIFIYSKKSLFSDPKETEKHPLSCWTLDILSQ
                    FIKCNLKGKISMDVHQSAVDCIHRIVSYKKKTLNRLPYKWSELWSALFSLISMVSTLN
                    ISNDKIQLNQSIHVGISTVNIFNLFINYGDSFLPSPNDYDELFYEIIRSGQVIENFYQ
                    FILNQTTTGGQSVTSPSAAASAFNPTTPTPLSPQQLPTITTTTTTTITSNNESNNPLLN
                    QLLNIRSIVQHFTSKLEEWSANNAEVALTAPQVSKIIKDNYDTLRLKLQENLDQFEPY
                    VENPKEVLFFKHIIKELINDLRKQINVTNII"
     CDS            complement(9007. .10431)
                    /note="GeneID exon scores (in order of location ranges):
                    117.93 - GSCJ_ID dd_03414"
                    /codon_start=1
                    /product="similar to Plasmodium falciparum. Hypothetical
                    protein"
                    /protein_id="AAS38634.1"
                    /db_xref="GI:42733694"
                    /translation="MESYSPSIQKLCVFLKPESDISNTGISYLNNFLADMSNAIIVES
                    IHSSKSHKHKTILIEDIESALNVVFPSEIADSIINQSKGRIQSTSLKNLMQVKTFNNG
                    GNNNNNNEQDNDLQEKEQNELSDDSDVEGFEIIPNNNNIYSHDVSLCFPIDLVCRVTL
                    DYICRKKVIHQNPSDFYKVIEQQQSNHVIYTYISYALEVITTKILIKISSSTDLIDED
                    IIRKVIQDNHHLNSVRHKNIQASNIAGKLLDSSSSSNSNLQTSLSSSFIFSSSSNNTT
                    TFTQQEQQEQFNDDKISGDEIEFEINNSKYLEKEDKKDQLIQKDEEIILKTFNDNNNN
                    NNNNNNNNNNNNNNNNNNININNDNDNNNNNNKNNDNKNYDNNNNNNADKIEKLSPVN
                    NNVKNSEIEKNIEKLKNINNNNNVEINEALISIIKETCKNIPSSSFSSVKNLKRSPPS
                    LFKFKEPIDLVQVLAEGGIHMEAT"
     CDS            complement(11849. .12583)
                    /note="GeneID exon scores (in order of location ranges):
                    50.22 - GSCJ_ID dd_02577"
                    /codon_start=1
                    /product="hypothetical protein"
                    /protein_id="AAS38635.1"
                    /db_xref="GI:42733695"
                    /translation="MKLLFGLILLIVNYCYLINANKLFVNFVGTYNSESCSESPIGTG
                    FSMALGQCVQMQGIFSYVTSSSPEATIWMATLGGDGNSFTLKQFHSSDKLCTEDPINT
                    VKYNKFDTCVQQPSFLTSNNSLLANNNNNNNNNNNNNNNNNNNNNNNLVYSKLTLSQNAPS
                    YAPNSVIIARYNSSDSVCNYNNQNYGISYSTGSVIYNENGVSVYYVCDQSNQSILFHC
                    QNDQCVQSFTDTSCDMPPNDLQIYCN"
     CDS            complement(13151. .13840)
                    /note="GeneID exon scores (in order of location ranges):
                    57.23 - GSCJ_ID dd_02576"
                    /codon_start=1
                    /product="hypothetical protein"
                    /protein_id="AAS38636.1"
                    /db_xref="GI:42733696"
                    /translation="MKLLLTLILIIVNYCCFINANKLFVNFVGSYSSNTCDGNPSGIG
                    FSLAVGQCLPIEGVLSYVINDMPPTTNWFITMGGNGNSFTLKQYDSQDQWCKQDPINT
                    LQFDNFDTCVEQPAFINVNNSLLADTPVVYSKLTLSQNVPTYAPDSVIIGQYNTNTNV
                    CTYNDQQFGVSYTSGTVIYNENGISIHVLCNQVNQSMMFKCKESKCGETFIDTYCQMS
                    GDVQIYCNSNN"
     CDS            join(20802. .21095,21199. .21978)
                    /note="GeneID exon scores (in order of location ranges):
                    23.32, 97.21 - GSCJ_ID dd_01856"
                    /codon_start=1
                    /product="similar to Dictyostelium discoideum (Slime
                    mold). Guanine nucleotide-binding protein alpha-2 subunit
                    (G alpha 2)"
                    /protein_id="AAS38637.1"
                    /db_xref="GI:42733697"
                    /translation="MGICASSMEGEKTNTDINLSIEKERKKKHNEVKLLLLGAGESGK
                    STISKQMKIIHQSGYSNEERKEFKPIITRNILDNMRVLLDGMGRLGMTIDPSNSDAAV
                    MIKELTSLQASIVTDCWGELNEDQGKKIKALWTDPGVKQAMRRANEFSTLPDSAPYFF
                    DSIDRMTSPVYIPTDQDILHTRVMTRGVHETNFEIGKIKFRLVDVGGQRSERKKWLSC
                    FDDVTAVVFCVALSEYDLLLYEDNSTNRMLESLRVFSDVCNSWFVNTPIILFLNKSDL
                    FREKIKHVDLSETFPEYKGGRDYERASNYIKERFWQINKTEQKAIYSHITCATDTNNI
                    RVVFEAVKDIIFTQCVMKAGLYS"
     CDS            join(22686. .22743,22820. .22926,23004. .23139,23232. .23558,
                    23639. .27180)
                    /note="GeneID exon scores (in order of location ranges):
                    -4.93, 3.57, -1.17, 18.69, 174.64 - GSCJ_ID dd_01857"
                    /codon_start=1
                    /product="hypothetical protein"
                    /protein_id="AAS38638.1"
                    /db_xref="GI:42733698"
                    /translation="MNRNIHLYIFVSILLSLINGIKSIKVGNGDFSVIDITDSFDDLV
                    TLYNDQYSIGGLDENIFWLKVINTVQIGENQKFNLLLTLDNSTTIQFDDFFTLTCKER
                    PVNIERELVGQKLFIDKQYKGRYGILVHNIGLEQYLSVPGLENNFISNTFTQGYYDDS
                    IFYTDVSKNYTIFKPVIKRSASVISQLSIDTIFGSVSTSTKLYLTSFLNSTYFKNPTI
                    APLQHDNNCSSLVYDQCSLLFVVQADFQRPGIYFLMNMDSNSNLPELVYGNKSYGHWL
                    FTIKNIAERQGSYIYPISSFDSTLGFSTIFSRNISYFGPTGVNLPVIQQSYFGTQIQN
                    SIFISFSKFVIPTTVNYGDIFIDNGFNTLSHFKYPYFINTINAQGGATNTLWVPSSPY
                    QKSIALEITLQFYFIKYPSPILVGGISGIASFKESIPPVFENIEYFDSVSDDCIVRIS
                    ASDEGGSCIQSIAISFIGAYEPSKVQSIFRITNRNLIEGDCFKGIYQISIPKYSIRRN
                    YFIEMNDYANNLVKLYPNNFYSGKRYLGSFPYQVLTLEDIESIGFRNNNIDLSNSNSS
                    NVLYIKLKDQSISNLVFQFIPKLGIDTYEPNNKISGTPRFFKSKWNSASSIYEVSFDL
                    PARLPTGPISYILNPLEIDSSILYQYFGSPSELNVYSQIADIMGPIALSYNTIGNTTI
                    PSINEPPNTIGFQILFGDYLNGISQVNITVTSDKTYKQMFKNFTFTEPSLGKWIQFTW
                    EIKYGDIPQTFSISELSAIDSNGYITEFPNPTKVNPLMLLVELYPSFSIQTSGETPTD
                    VTPPYMQVYHVNQLNIENYLDRTLEINFTVYDGESGIYIEEPPTIYFINDNVEIISKQ
                    TSYISGTSSGLGFFENAIFGVNITLPYLFGYPNGFLTQIHGIYDNKFNVLGATPDIIN
                    QYNRGNSFINVTIGQSQPFIESFVMDIDGDLLTVNGLKFGILPTGKLRINGAIENLNI
                    SLPFLNNTYIVFDLKQYRNINDTNCQIYITNQLGFSSNEINFVLKAHTNVTMPPTTTP
                    HPTSNPTATPSLTANPTPLPTNSPSQCKNDCGGIAQGQCSNLGCICISPYIGIDCSSK
```

```
                     VIVDIDPIVNNSQPEVITPYQHENSGNKFSTLVSIYSLRELGFDGKVENDYIFEMWNY
                     QYYNNDNHTYFTSIKNKNDLSKRTNITVSIKMYKQLTTIQFAGENITIQPSSIKYFIS
                     LSDYQFKSNLNTLDLVMYTSIETEKIDDACFSSEFGDSSSQLSSQYLKIQINDVSLYG
                     EFIKRAIVDTHIVGVLNKVRGNNINSTSNGSSSSQYISIQIPNYKTSVELDPNFSVLL
                     ESSKIENSTCNKNQSGLTKTQLIGIIIGSTAGFILILIIVISIVYKKSTTLRININRA
                     KKFRMKKLSFK"
     CDS             complement(join(27340. .29056,29330. .29490))
                     /note="GeneID exon scores (in order of location ranges):

Alignment Scores:
Pred. No.:                43.1          Length:          88549
Score:                    73.00         Matches:         12
Percent Similarity:       85.71%        Conservative:    0
Best Local Similarity:    85.71%        Mismatches:      0
Query Match:              86.90%        Indels:          2
DB:                       2             Gaps:            1

US-09-858-332G-15 (1-12) x AC116924 (1-88549)

Qy          1 HisAsn------HisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||      ||||||||||||||||||||||||||||||
Db      46092 CACAATATGGTTCACAATCACAATCACAATCACAATCACAAT 46051

RESULT 276
NCB23B10/c
LOCUS       NCB23B10               99967 bp    DNA     linear   PLN 17-APR-2005
DEFINITION  Neurospora crassa DNA linkage group II BAC contig B23B10.
ACCESSION   BX284752
VERSION     BX284752.1  GI:28881372
KEYWORDS    .
SOURCE      Neurospora crassa
  ORGANISM  Neurospora crassa
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
REFERENCE   1
  AUTHORS   Schulte,U., Aign,V., Hoheisel,J., Brandt,P., Fartmann,B.,
            Holland,R., Nyakatura,G., Mewes,H.W. and Mannhaupt,G.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 99967)
  AUTHORS   German Neurospora genome,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAR-2003) MIPS, Institut fuer Bioinformatik,
            GSF-Forschungszentrum fuer Umwelt und Gesundheit, GmbH,
            Ingolstaedter Landstrasse 1, D-85764 Neuherberg, FRG, E-mail:
            G.Mannhaupt@gsf.de Project Coordinator: Ulrich Schulte, Institute
            of Biochemistry, Heinrich-Heine-University, D-40225 Duesseldorf,
            E-mail: ulrich-schulte@uni-duesseldorf.de
COMMENT     this contig is an assembly of BAC 23B10 from 1 to: 41462, strain
            OR74A, cosmid 1K2 from 41463 to: 69790, cosmid library at DKFZ,
            strain OR74A and cosmid 39C11 from 69791 to 99967, cosmid library
            pLORIST6Xh, strain 74-OR-23-1A; BAC clone 23B10 and cosmid clone
            39C11 is available at the Fungal Genetic Stock Center,
            http://www.fgsc.net cosmid clone 1k2 is available from J. Hoheisel
            at DKFZ, e-mail:
            J.Hoheisel@dkfz-heidelberg.de
            Sequencing was performed by MWG Biotech AG, Ebersberg, Germany,
            http://www.mwgdna.com
            Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of linkage groups II
            and V can be viewed at: http://mips.gsf.de/proj/neurospora.
FEATURES             Location/Qualifiers
     source          1. .99967
                     /organism="Neurospora crassa"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5141"
                     /chromosome="6"
     gene            4533. .8183
                     /gene="B23B10.010"
     gene            complement(4533. .8183)
                     /gene="B23B10.010"
     CDS             complement(4533. .8183)
                     /gene="B23B10.010"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAD70414.1"
                     /db_xref="GI:28881373"
                     /db_xref="UniProt/TrEMBL:Q872Z6"
                     /translation="MYPHRRAPGDQGEHVIWRSNTADIPDDVIDPQLLAVSNQSDAAH
                     QPSLPQQPSQLLAETSQSNTVEQTSVYLHPPLDPVHQYNPQVELQPNRHVSLPPRDGS
                     IYQQQQLDFHHPNAFSNPHVNPQVISPSNSQPNPQYSFSPQSHSPFRDGSIHHHQPFN
                     PSSLAYSPPQDGRAFFQQAASSPSLPRHSRTNRMSQGDVRASPLIRTPHRNQDRSVRA
                     NVPSPAKSKSRAFSRGFGQGPMSEADEGPFRFQDEPDSSPLTSLTATPARSSPDDGE
                     MEMDLDDWESGREAEVGIPSSSPPQNNVTPTRPRQRQEEDLQGREGQYQRPRRYNIDL
                     LLESVAITPVADMEIEEGVDEAASHADVESASHDPSLQRPESQENARPRAASLEIVTP
                     PRGSILRRFGPGTGLSFPSGVPYSTRYSDYTPPPEDFWRPLFAARPVPPGGYLAVERA
                     ADERTQTDNVSGPQDGESFRHNQHLNSQTGGVRHHMRTAIPGPSPTASTAPNNPQPTT
                     SRHGPASLPQGASQDTSVNDVVEVTSTDTILTHHRGPQRHQGHLRPLAPKPPIEDKDT
                     SPQPPHHTAQHPTQGNDQTQTQQPRFTNMAQFQPLLALPPPIQHPLLLSPFPPRIPGT
                     LTPALLSLAINPPRSLQLGTTDSLQPYLRACGLPVLYNDVYPRALPISVRREPFNRTV
                     IVEERYLAVIGLLYPRAEFIVSRIAEGVGNQKVEPNSGVMMDKWHVYKAVRMSGEEMQ
                     KLGVCVWSEKVCGEWQPVVEEREGREVKRVKMMDVVVDEEEEDETDEESDDIVVVSGR
                     ARASEKEGKENVRMTVGEERDEDCEDMEIVDAAAEGVQTSKTTNGKEKAKEVQVADKE
                     DKGKEILQEINDDDNTAAPPEPTSSSASAPLRRSTRVQSRQASVEEPATDKSENTTTA
                     PSNPSGKTLITRKLRAIQLDETGATITPNPNPTRPTKTPKPRLSTAITTTTTTTTKPP
                     APTVPFSNPSTVPSRSSNPTPSHRTSPTHYTVTLRQRDYPLGGVARIQCTPRARQVIE
                     SIQASFGVQHPYEVWRHKDYTVHDPDTGEVTQRYIKVLRGYAGDENTDDGLMRLFLDE
                     MDQRIKLKDTENKRNEKGKKKKDGQEEEEGKDMAALEDRVAWKERYLALKRRGKMAVA
                     RAKERKRDEEDMDMDYDEWLEERHEQEEELIEKWFAERSGLVEEEELARKTVEFALFE
                     YGGGDGLALGKP"
     exon            complement(4533. .8183)
                     /gene="B23B10.010"
                     /number=1
     tRNA            8734. .8806
                     /product="tRNA-Val"
                     /note="tRNA predict as a tRNA- Val : anticodon cac"
     gene            9902. .10129
                     /gene="B23B10.020"
     gene            complement(9902. .10129)
                     /gene="B23B10.020"
     CDS             complement(9902. .10129)
                     /gene="B23B10.020"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAD70415.1"
                     /db_xref="GI:28881374"
                     /db_xref="UniProt/TrEMBL:Q872Z5"
                     /translation="MSSSNTDTSTTTTSGDNTSSTSTTTQTNSSTTSSSSSSSTSTTS
                     KTTGSSGYSYIERLRIEGQAQYWLGPEKKSS"
     exon            complement(9902. .10129)
                     /gene="B23B10.020"
                     /number=1
     gene            13672. .15211
                     /gene="B23B10.030"
     CDS             join(13672. .14823,14894. .15211)
                     /gene="B23B10.030"
                     /note="similarity to monocarboxylate transporter (mct3),
                     homo sapiens, TREMBL:U81800_1
                     Contains Multicopper oxidases signatures
                     [GPYVLFTFGTFVGFIGLYVML]
                     contains EST gb:AA898910, BG279307"
                     /codon_start=1
                     /product="related to monocarboxylate transporter"
                     /protein_id="CAD70416.1"
                     /db_xref="GI:28881375"
                     /db_xref="GOA:Q872Z4"
                     /db_xref="InterPro:IPR002355"
                     /db_xref="InterPro:IPR007114"
                     /db_xref="InterPro:IPR011701"
                     /db_xref="UniProt/TrEMBL:Q872Z4"
                     /translation="MAPTTFPDRTTHGPSTDAEGNLSDQSTATGISSPSRASSEKSSP
                     HVTAAPPLTGAAASSVAPDAEKQQPPPTTAAWNAGPPPPNGGLAAWLQVVSVFFIFFN
                     TWGLLNTFGIFQTYYESSESIFQPPGSVSSSNISWIGSIQSFMVLLGGLVSGPLYDRG
                     YLKYLIYVGAFGVVFGHMMLSLVSEFWQALLAQGFVVGLGAGMMFTPGVTVLQSYFST
                     KIGLATGLAAAGSSFGGVIYPIVFYRLIDRIGYGWSVRVIGFIALATLMVPVVFLKQR
                     TKSGKVRALVDVSAFTDGPYVLFTFGTFVGFIGLYVMLFYLSFFGLDTGATDPAMAFY
                     IVPILNAASMFGRTLPNWLSDKVGPFNILIPGALVCGALTLAMMGVKGLGAIIVVAIL
```

```
                FGFFSGVFIALPTVCYIALTKDKSRLGSRIGLGYAFLGLATLIGGPGTGAILQQTGHN
                GHNNWTGIWVFGGVMLLAASGIFMVLRSWKFGFKLTTKA"
     exon       13672. .14823
                /gene="B23B10.030"
                /number=1
     intron     14824. .14893
                /gene="B23B10.030"
                /number=1
     exon       14894. .15211
                /gene="B23B10.030"
                /number=2
     gene       19788. .21920
                /gene="B23B10.040"
     gene       complement(join(19788. .21110,21174. .21622,21848. .21920))
                /gene="B23B10.040"
     CDS        complement(join(19788. .21110,21174. .21622,21848. .21920))
                /gene="B23B10.040"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="CAD70417.1"
                /db_xref="GI:28881376"
                /db_xref="GOA:Q872Z3"
                /db_xref="InterPro:IPR002156"
                /db_xref="UniProt/TrEMBL:Q872Z3"
                /translation="MDKGQPPDHPASEQEPPAKRHKPLNSYFPQPFRGKIFVQNERPR
                EEEAIAAAREAIASAKRSTENDFTIAFFTDASVEEKTPDLPGGYSVVFDNFFPGTHHY
                RERLGMAWLMPNMITSTIGEAVGILQALQVAIHHLRAIAISPADTASARRVTLCIFSD
                AHQVIRLIKCQNLNVGDIPGFDVFLELHWVKGHAGVEENDQADLIARRARPIEDGRDI
                FVASGADVDATALPPAVHLLFLDDIAAEKARQRPPRPLGKQNDRRKRNGKRAHRRRDE
                NVPSSSLHTPNDDMTYSPPPWASDSFRFSTNSLEMPQYPATRPFSFASEEHPRVSEKP
                TLETEEQSPMPKQYTPLDVNNHEKKPAVEIEALERDSTLEQGVAVIQPAVIQQAIVEK
                EAAVLNINKPTIGLDSSANEQALTLLQHEEQEELMTDLEVQVPEKNSVLEHIVSDSLA
                SVPQQVVEEDKHSVKREVQKPAILLSTGMEVAENKQDATPHSVNEQHEPSTKAEDQKF
                IIPEAIEDTEPSPDSEGSTAVHQPAEEHNDSALEPLLGDQALIDLQLEGEYVEWCQRR
                LQDVLVSVYNHEGGSTGETDTSEEVISGEETQAAVGRPLAEEVEATLA"
     exon       complement(19788. .21110)
                /gene="B23B10.040"
                /number=1
     intron     complement(21111. .21173)
                /gene="B23B10.040"
                /number=1
     exon       complement(21174. .21622)
                /gene="B23B10.040"
                /number=2
     intron     complement(21623. .21847)
                /gene="B23B10.040"
                /number=2
     exon       complement(21848. .21920)
                /gene="B23B10.040"
                /number=3
     gene       22992. .23446
                /gene="B23B10.050"
     gene       complement(join(22992. .23008,23188. .23446))
                /gene="B23B10.050"
     CDS        complement(join(22992. .23008,23188. .23446))
                /gene="B23B10.050"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="CAD70418.1"
                /db_xref="GI:28881377"
                /db_xref="UniProt/TrEMBL:Q872Z2"
                /translation="MSNIFFWLPMEDQAKVYALMWPGNWKKKKKKKKKKKKKKKSSA
                SLASQSMGASGSGQKEKASSDGGSGGSKDGKMETEVETVEYKESWAR"
     exon       complement(22992. .23008)
                /gene="B23B10.050"
                /number=1
     intron     complement(23009. .23187)
                /gene="B23B10.050"
                /number=1
     exon       complement(23188. .23446)
                /gene="B23B10.050"
                /number=2
     repeat_region 23323. .23370
                /note="48 bp tcttcttcttcttcttcttcttcttcttcttcttct tandem
                repeat"
     gene       24423. .26455
                /gene="B23B10.060"
     CDS        join(24423. .24497,24557. .26455)
                /gene="B23B10.060"

Alignment Scores:
Pred. No.:              48.4          Length:         99967
Score:                  73.00         Matches:        10
Percent Similarity:     100.00%       Conservative:   1
Best Local Similarity:  90.91%        Mismatches:     0
Query Match:            86.90%        Indels:         0
DB:                     15            Gaps:           0

US-09-858-332G-15 (1-12) x NCB23B10 (1-99967)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              |||:::|||||||||||||||||||||||||||
Db       5310 CACAGCCATAACCACAACCACAACCACAACCAC 5278

RESULT 277
AE003426_2/c
WPCOMMENT
  Sequence split into 4 fragments  LOCUS AE003426 Accession AE003426
     Fragment Name      Begin     End
     AE003426_0             1     110000
     AE003426_1         100001    210000
     AE003426_2         200001    310000
     AE003426_3         300001    396604
Continuation (3 of 4) of AE003426 from base 200001 (AE003426 Drosophila melanogaster chro

Alignment Scores:
Pred. No.:              53.1          Length:         110000
Score:                  73.00         Matches:        10
Percent Similarity:     100.00%       Conservative:   1
Best Local Similarity:  90.91%        Mismatches:     0
Query Match:            86.90%        Indels:         0
DB:                     2             Gaps:           0

US-09-858-332G-15 (1-12) x AE003426_2 (1-110000)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              |||||||||||||||||||||||:::|||
Db      45221 CACAATCACAATCACAATCACAATCACGATCAC 45189

RESULT 278
AC104144
LOCUS       AC104144              179069 bp    DNA     linear   HTG 18-MAY-2002
DEFINITION  Drosophila melanogaster chromosome X clone RP98-23J3, WORKING DRAFT
            SEQUENCE, 2 unordered pieces.
ACCESSION   AC104144
VERSION     AC104144.4  GI:20898989
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 179069)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
```

```
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 179069)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 179069)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 17, 2002 this sequence version replaced gi:17940809.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: DROB
                Center clone name: RP98-23J3
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Sequencing vector: M13;
                Chemistry: Dye-terminator Big Dye: 99% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 177843 bases at least Q40
                Consensus quality: 178008 bases at least Q30
                Consensus quality: 178254 bases at least Q20
                Estimated insert size: 175210; sum-of-contigs estimation
                Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    73824: contig of 73824 bp in length
            *    73825    73924: gap of unknown length
            *    73925   179069: contig of 105145 bp in length.
FEATURES                Location/Qualifiers
     source             1. .179069
                        /organism="Drosophila melanogaster"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:7227"
                        /chromosome="X"
                        /clone="RP98-23J3"
     gap                73825. .73924
                        /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:                  84.8          Length:          179069
Score:                      73.00         Matches:         10
Percent Similarity:         100.00%       Conservative:    1
Best Local Similarity:      90.91%        Mismatches:      0
Query Match:                86.90%        Indels:          0
DB:                         14            Gaps:            0

US-09-858-332G-15 (1-12) x AC104144 (1-179069)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              ||||||||||||||||||||||||||:::|||
Db      70150 CACAATCACAATCACAATCACAATCACGATCAC 70182

RESULT 279
AC116543
LOCUS       AC116543               182387 bp    DNA     linear   INV 22-MAR-2003
DEFINITION  Drosophila melanogaster X BAC RP98-1A14 (Roswell Park Cancer
            Institute Drosophila BAC Library) complete sequence.
ACCESSION   AC116543
VERSION     AC116543.6  GI:29150331
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 182387)
  AUTHORS   Muzny,D., Scherer,S., Adams,M.D., Holt,R.A., Evans,C.A.,
            Gocayne,J.D., Tabor,P., Williamson,A., Homsi,F.H.,
            Dugan-Rocha,S.D., Sodergren,E.S., Hodgson,A.H., Chen,R.C.,
            Ayele,M., Scott,G.S., Worley,K.W., Amamatides,P.G., Brandon,R.C.,
            Rogers,Y., An,H., Baldwin,D., Beeson,K.Y., Brown,M., Buhay,C.,
            Busam,D.A., Center,A., Chen,G., Chen,Z., Clerc-Blankenburg,K.,
            Davenport,L.B., Dietz,S.M., Ding,Y., Dodson,K., Doup,L.E.,
            Draper,H., Emery-Cohen,A., Ferriera,S., Garg,N.D.S., Houck,J.,
            Hostin,D., Howland,T.J., Hume,J., Ibegwam,C., Jalali,M., Kovar,C.,
            Liu,W., Mattei,B., McIntosh,T.C., Morgan,M., Moy,M., Murphy,B.,
            Nelson,K.A., Ndassa,Y., Nguyen,N., Perez,L., Pittman,G.S., Puri,V.,
            Scheeler,F., Shen,H., Strong,R., Tector,C., Wang,Q., Williams,S.M.,
            Xiang,J., Zaveri,J.S., Zhou,J., Zorrilla,S., Smith,H.O.,
            Wheeler,D., Weinstock,G., Gibbs,R. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 182387)
  AUTHORS   Worley,K.C., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
```

Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 182387)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (29-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 4 (bases 1 to 182387)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (16-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 5 (bases 1 to 182387)
AUTHORS BCM-HGSC.
TITLE Direct Submission
JOURNAL Submitted (22-MAR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Mar 22, 2003 this sequence version replaced gi:22902011.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:
http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

FEATURES Location/Qualifiers
source 1. .182387
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/chromosome="X"
/clone="RP98-1A14"
misc_feature 123441. .123464
/note="homopolymeric run"
/function="size variation in subclone"
misc_feature 166110. .169176
/note="size
3067 bp; confirmed by restriction digest"
/function="unresolved tandem repeat"
misc_feature 168452. .168456
/function="low quality"
misc_feature 181287. .181298
/note="homopolymeric run - Baylor 12 As/Celera 11 As"
/function="polymorphic site"
ORIGIN

Alignment Scores:

Pred. No.:	86.3	Length:	182387
Score:	73.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	90.91%	Mismatches:	0
Query Match:	86.90%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-15 (1-12) x AC116543 (1-182387)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              ||||||||||||||||||||||||||:::|||
Db      51386 CACAATCACAATCACAATCACAATCACGATCAC 51418
```

RESULT 280
AC097439
LOCUS AC097439 221040 bp DNA linear HTG 10-MAY-2003
DEFINITION Rattus norvegicus clone CH230-150M9, WORKING DRAFT SEQUENCE, 2 unordered pieces.
ACCESSION AC097439
VERSION AC097439.6 GI:30521188
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 221040)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,

```
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 221040)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 221040)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:23617376.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GIOL
                Center clone name: CH230-150M9
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 214277 bases at least Q40
                Consensus quality: 215879 bases at least Q30
                Consensus quality: 217162 bases at least Q20
                Estimated insert size: 225376; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1   219392: contig of 219392 bp in length
            *    219393   219492: gap of unknown length
            *    219493   221040: contig of 1548 bp in length.
FEATURES              Location/Qualifiers
     source           1. .221040
                      /organism="Rattus norvegicus"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:10116"
                      /clone="CH230-150M9"
     misc_feature     106906. .118481
                      /note="clone_boundary
                      clone_end:T7
                      site:EcoRI
                      end_sequence:BZ105133"
     gap              219393. .219492
                      /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:                104           Length:          221040
Score:                    73.00         Matches:         10
Percent Similarity:       100.00%       Conservative:    2
Best Local Similarity:    83.33%        Mismatches:      0
Query Match:              86.90%        Indels:          0
DB:                       14            Gaps:            0

US-09-858-332G-15 (1-12) x AC097439 (1-221040)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||||||:::|||||||||||||||||||||
Db     219945 CACGACCACAACTACAACCACAACCACAACCACAAC 219980

RESULT 281
AC098906/c
LOCUS       AC098906               221938 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-150C15, WORKING DRAFT SEQUENCE, 3
            unordered pieces.
ACCESSION   AC098906
VERSION     AC098906.7  GI:30521190
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 221938)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
```

Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE Direct Submission
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 221938)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 221938)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On May 10, 2003 this sequence version replaced gi:23617294.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GION
Center clone name: CH230-150C15
---------- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 212507 bases at least Q40
Consensus quality: 214312 bases at least Q30
Consensus quality: 215742 bases at least Q20
Estimated insert size: 225546; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 38090: contig of 38090 bp in length
* 38091 38190: gap of unknown length
* 38191 139254: contig of 101064 bp in length
* 139255 139354: gap of unknown length
* 139355 221938: contig of 82584 bp in length.

FEATURES Location/Qualifiers
source 1. .221938
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-150C15"
misc_feature complement(4481. .5286)
/note="clone_boundary
clone_end:T7
site:EcoRI
end_sequence:RWBAU20TJC"
gap 38091. .38190
/estimated_length=unknown
misc_feature 38191. .40787
/note="wgs_contig"
gap 139255. .139354
/estimated_length=unknown

ORIGIN

Alignment Scores:

Pred. No.:	104	Length:	221938
Score:	73.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	83.33%	Mismatches:	0
Query Match:	86.90%	Indels:	0
DB:	14	Gaps:	0

US-09-858-332G-15 (1-12) x AC098906 (1-221938)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::||||||:::|||||||||||||||||||||
Db      79765 CACGACCACAACTACAACCACAACCACAACCACAAC 79730
```

RESULT 282
AC113650
LOCUS AC113650 248646 bp DNA linear HTG 21-SEP-2002
DEFINITION Rattus norvegicus clone CH230-45C16, *** SEQUENCING IN PROGRESS ***.
ACCESSION AC113650
VERSION AC113650.4 GI:23266026
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 248646)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,

Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 248646)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (05-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 248646)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (21-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Sep 21, 2002 this sequence version replaced gi:21743105.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequening reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GRPM
Center clone name: CH230-45C16

---------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 219151 bases at least Q40
Consensus quality: 222881 bases at least Q30
Consensus quality: 225214 bases at least Q20
Estimated insert size: 242971; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 248646: contig of 248646 bp in length.

FEATURES Location/Qualifiers
source 1. .248646
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-45C16"
misc_feature 1. .2100
/note="wgs_end_extension
clone_end:Sp6"
misc_feature 8813. .9624
/note="clone_boundary
clone_end:Sp6
site:EcoRI
end_sequence:BH296751"
misc_feature 46244. .47822
/note="wgs_contig"
misc_feature complement(244139. .244864)
/note="clone_boundary
clone_end:T7
site:EcoRI
end_sequence:BH296749"
misc_feature 246213. .248646
/note="wgs_end_extension
clone_end:T7"
ORIGIN

Alignment Scores:

Pred. No.:	116	Length:	248646
Score:	73.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	83.33%	Mismatches:	0
Query Match:	86.90%	Indels:	0
DB:	14	Gaps:	0

US-09-858-332G-15 (1-12) x AC113650 (1-248646)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::||||||:::||||||||||||||||||||||
Db     106117 CACGACCACAACTACAACCACAACCACAACCACAAC 106152
```

RESULT 283
AC132543/c
LOCUS AC132543 254668 bp DNA linear HTG 19-NOV-2002
DEFINITION Rattus norvegicus clone CH230-96C8, *** SEQUENCING IN PROGRESS ***, 2 unordered pieces.
ACCESSION AC132543
VERSION AC132543.3 GI:25072633
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 254668)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 254668)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 254668)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Nov 19, 2002 this sequence version replaced gi:23683155.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
-------------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-------------- Project Information
Center project name: GRZW
Center clone name: CH230-96C8
-------------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 237238 bases at least Q40
Consensus quality: 241030 bases at least Q30
Consensus quality: 243348 bases at least Q20
Estimated insert size: 246228; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 136730: contig of 136730 bp in length
* 136731 136830: gap of unknown length
* 136831 254668: contig of 117838 bp in length.

FEATURES Location/Qualifiers
source 1. .254668
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-96C8"
misc_feature 1. .2553
/note="wgs_end_extension
clone_end:Sp6"
misc_feature 2486. .3234
/note="clone_boundary
clone_end:Sp6
site:
end_sequence:BH363340"
misc_feature 6656. .65620
/note="clone_boundary
clone_end:T7
site:
end_sequence:BH363339"
gap 136731. .136830
/estimated_length=unknown
misc_feature 200078. .201080
/note="wgs_end_extension
clone_end:T7"

ORIGIN

Alignment Scores:

Pred. No.:	119	Length:	254668
Score:	73.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	83.33%	Mismatches:	0
Query Match:	86.90%	Indels:	0
DB:	14	Gaps:	0

US-09-858-332G-15 (1-12) x AC132543 (1-254668)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||||||:::|||||||||||||||||||||
Db      76869 CACGACCACAACTACAACCACAACCACAACCACAAC 76834
```

```
RESULT 284
AC117075/c
LOCUS       AC117075              254733 bp    DNA     linear   INV 12-MAR-2003
DEFINITION  Dictyostelium discoideum chromosome 2 map 5201047-5455781 strain
            AX4, complete sequence.
ACCESSION   AC117075 AC115615 AC117177
VERSION     AC117075.2  GI:28828046
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 254733)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 254733)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-APR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 254733)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 254733)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On or before Mar 4, 2003 this sequence version replaced
            gi:20069690, gi:19570144, gi:20066242.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES              Location/Qualifiers
     source           1. .254733
                      /organism="Dictyostelium discoideum"
                      /mol_type="genomic DNA"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /chromosome="2"
                      /map="5201047-5455781"
     CDS              complement(join(<124. .619,719. .1068,1167. .1457))
                      /note="GeneID exon scores (in order of location ranges):
                      30.56, 24.41, 16.56 - GSCJ_ID dd_01823"
                      /codon_start=1
                      /product="hypothetical protein"
                      /protein_id="AAM09355.2"
                      /db_xref="GI:28828135"
                      /translation="MGIEGLNLFMGKQFGCVEKVKEIPTDHVYIDLNNYLHKSVKRGD
                      NSKLNTEVNVFRILKSFIDGILRKVRVKHSVFFGIDGPGPRSKMILQRERRLKNGNID
                      KLKYFINRNKEQQQQQQQQQQSPQLHDYDYLNENDLKSFDSSFSTLEFTPGTTFMGK
                      LKDFLIFYTKNKLYFAKKIFISASDRIGEGEFKIFEQILNSNYPINDSFTIVSSDSDI
                      LLFSLLSKYKNVYIYNKDSEEIIKIDKIRDKIYQQCYKKKQKKQKQQKKSGGEVEGEG
                      EGIATEEKEEGVQIEEEEEEEEEEEEEDISKRRQAIVDFVMLTFLMGTDHLPKVSSYNI
                      GSAWSEYCKIKKPLYNEETGLINLDLLFRLIGKSSQPRLFYRNRF"
     CDS              complement(join(1796. .3073,3249. .3563))
                      /note="GeneID exon scores (in order of location ranges):
                      87.30, 9.45 - GSCJ_ID dd_01824"
                      /codon_start=1
                      /product="hypothetical protein"
                      /protein_id="AAO50731.1"
                      /db_xref="GI:28828047"
                      /translation="MKLGRDIPIYKLIIIKNESIENNLKEISFSDNFKCQICEGLLIS
                      SLIPNRMKALQCINGHCFCLTCWESILEIKSECPTCRIQIQSMNTLSNNLFIIKSISE
                      SIKIHCPNYLNFDNSNNFNGCKEIITIDEIDRHESKCEFRFIKCSINNQCNEIIRFNE
                      RDKHESQCDFIIQKCTHCDESVQMKQMQGHILFECLKVLITCQHCSLQFTRLKLQNHI
                      ENHCKEIKIECPFKSLKIIYNNNNNDNDNNDSDENNSNQSLSSSSLSSSIIPLSPCNE
                      LMKRSELSKHFIRNHQYHNELVSIVLKKQENKIKKLETIIQQNNVNQNFEISIIKHSN
                      SLTRDSFYKENKKKDKLIQDLMKRVLILEHQQQQNQIQNQIQNQIQKLSNKSIFTTTT
                      TTTSTSDKNNNNNNNNNNNNNNNEDEEDDDNIKNNDNQGNLKSLREMFNTISNFKDD
                      ENEQQQIQQQQQLPFTSFGPLSVQQQNNPTKQFNQLSQPQTQPQSQSQSQSLFGEWSP
                      ITINQNQNTPSNPFSIFSGTSL"
     CDS              complement(join(4148. .4536,4589. .4883,4975. .5261,
                      5357. .5513,5597. .5830,5946. .6098))
                      /note="GeneID exon scores (in order of location ranges):
                      33.35, 9.09, 28.75, 10.40, 9.66, 11.24 - GSCJ_ID dd_01825"
                      /codon_start=1
                      /product="similar to Mus musculus (Mouse). 0 day neonate
                      eyeball cDNA, RIKEN full-length enriched library,
                      clone:E130103I15 product:polycystic kidney disease 2, full
                      insert sequence"
                      /protein_id="AAO50732.1"
                      /db_xref="GI:28828048"
                      /translation="MARSGYYIDIPVSDIKTGVESLIDNGFFDIQTRSVIISFSTLNL
                      NFQSRTSSGQHQDQLIPIISIRTYRIQLYMDALDKFRGFLEVSFFIFILYYFYYFVNE
                      ARIDFKLGRFKFYISNFKNIFDVINLTLFIASIALYLAFLGDNSRDLNINSQIQTTTD
                      YPLYLENLGQTALVLYQISAINILLMSFKTFRFLVVHRRLYILWIAISQSRLQLITFT
                      IMFCIMMLGFLFSGWLTFGTDIASFNGFVTSLGTLLQFIIGNPPDYEAMSYTNRALGP
                      IYYLLFTIFMFFILVNMFVAIISDSYQNISNTPEKKQKTKRYLLTRWERTMKSLLLLF
                      TPNSLDIYDLVLMFIDQKPGLIETDGIDPSRFKDEIMIIQPIQRARIKYLDQIQSAKN
                      TGVTFELKLMGQQVHKKDLDTWKKEAKIKEKKKFESIQQNINILNTKLDLLLQAMNVP
                      QQSIPQLINGSTGMPQSPIINNNNNNGGSIPSSSSNGSINGSNHINLSGGTNTP"
     CDS              complement(join(6215. .6478,6568. .6864))
                      /note="GeneID exon scores (in order of location ranges):
                      21.78, 25.12 - GSCJ_ID dd_03502"
                      /codon_start=1
                      /product="hypothetical protein"
                      /protein_id="AAO50733.1"
                      /db_xref="GI:28828049"
                      /translation="MNTLKRTVFSQRLEAVKALENKTREEKATLLAAVETQRNARLVK
                      DLIIHLVFFVVFVLIVVLQLNPQSLFIINTSLKNQILFSQDSQYLEINNPSDFITYLN
                      STFCESISDMSLENNDAMNKIIGNVIRIRTARVKPDSCSNNGFNLTCYSNNYDKDTKD
                      RSPFGPNDMYTYTSNSHGSYVFGMFF"
     CDS              join(8148. .8280,8374. .10196)
                      /note="GeneID exon scores (in order of location ranges):
                      7.51, 117.32 - GSCJ_ID dd_00414"
                      /codon_start=1
                      /product="similar to Suppressor of mutant AC40 subunit of
                      RNA polymerase I and III (high serine); Srp40p
                      [Saccharomyces cerevisiae]"
                      /protein_id="AAO50734.1"
                      /db_xref="GI:28828050"
                      /translation="MNKTLSNKKASSKTTNNSNNNNNNTINDPVLSLPTYSLEIYEED
                      YFNELILEPLSFLKSGDVVITTSPTKKKTSSVSASPSPIKKESNSNSNSNSNSNSNSN
                      SNNNNNDSDSSSDNDLKFEGNQDDSEDEDQTNSNYYCPFCKKKFQTEQTWNTHINSSK
                      HLKSIKDQKKKSTTTGTSNTPTKNNNNNSPSNVKSPSSTKTKSSSSSSSSSSSSSSSS
                      SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSTNSKDNEFLQDQIMQTQQAIKIKNSKP
                      NLACKNLFTIARSYCQKQYIRDSAKTLNLIIDILTISNNNNNNNTTPSITSTTTSSTS
                      NKSIVEYSLNNILNNATLLCKTHLYFGRLTRLFNQRLSEYHFLQCFNYLALINNDQLL
                      KYESSLVSSSPLNLYKIFIIQLDKFGKSKNKDTNTNTDTDTNTNTDTNKEIESIIQIN
                      KYINEIAGALCYCEVNFKSIILYILASCITFKFKLYKESLESNTRAFKLLSSLKQYHW
                      SIDLLLNNLLIKLKLNETIDIVDGDGDTEIQYQKLLLKEKEKEKEKEYQNIIIILIKY
                      CILIDDKIRLSAILGDYNFQKIISIYFKNNDYVDCLIKSSNQFLIENYKEIDQIKDHY
                      EHLIYSNKDFIFIFNQFDLKKKILLQK"
     CDS              complement(join(10448. .10915,11145. .11294))
                      /note="GeneID exon scores (in order of location ranges):
                      45.58, 25.53 - GSCJ_ID dd_00412"
                      /codon_start=1
                      /product="similar to Dictyostelium discoideum (Slime
                      mold). Ras-related protein Rap-1"
                      /protein_id="AAO50735.1"
                      /db_xref="GI:28828051"
                      /translation="MGKGNGKSKDSKGSEKSSTKTKFFVAVMGAGSVGKSALTVQFTQ
                      GVFIDKYDPTVEDTYTKTFELDGEQVCIEVLDTAGSEVLVAMRELYMKNAEGFILVYS
```

```
                ILVKSTFLEIKDIVEQLFRVKEEEEVPIVLVGNKIDLDAHREVSTNEGKQLANSYPNC
                DFWEASSKDRINVDNVFFSIVKRIKEKYKKEGAPVKEKKEKCIIM"
     CDS        12794. .13423
                /note="GeneID exon scores (in order of location ranges):
                35.19 - GSCJ_ID dd_00410"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAO50736.1"
                /db_xref="GI:28828052"
                /translation="MASIGTGKFKNEPTFGKNIKEILILLERSIAESEKQWVDLVSEN
                KNLVDVNAFRINKLFDNKIGLDVKDNKSFELLKKTVHEFKTERPVIDLVCKSISCLFY
                LSDLQLSSIITNLPNEIIAQLKDSTVPFAYRLENSSLKNVQIQKRHYLSSNRPIELVC
                CILYLPANIDIICKVVPKSKFLNETSISGCPFDLEMDSSFYGIKKKLSL"
     CDS        complement(join(13494. .13813,13915. .14097,14208. .14241))
                /note="GeneID exon scores (in order of location ranges):
                22.37, 12.24, 8.90 - GSCJ_ID dd_00408"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAO50737.1"
                /db_xref="GI:28828053"
                /translation="MSIYEEIMRAQVKQGNMGGRSGPLGYLLRENKQKEKKAVKVNTP
                KPPKPVVTQSPYTINLIEGTVTNATDNNGFSDPYLVIIAIKPDNSFSKILFKSNVCKK
                TLTPTWNQVENFFVKDNVTDIIVELWDHDVLSKNDFIGRSIISMNTVRSKGYDGTLEF
                FNGSEVVDGTVKLTIKRK"
     CDS        complement(join(14655. .14686,14848. .14916,14992. .15162,
                15268. .15310))
                /note="GeneID exon scores (in order of location ranges):
                1.29, 3.37, 1.88, 4.01 - GSCJ_ID dd_00406"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAO50738.1"
                /db_xref="GI:28828054"
                /translation="MSEICSTSCWQYAIDNPNCFPKIGFTKVCFMSGNYTCASSSGFT
                ILSDELSFINYGLYHPDEIIGKFIKIKGNEYKVTKATNSIINLSNSKNGRVTCLPKRR
                VV"
     CDS        complement(join(15876. .15900,16018. .17390))
                /note="GeneID exon scores (in order of location ranges):
                -0.56, 167.37 - GSCJ_ID dd_00405"
                /codon_start=1
                /product="similar to Physarum polycephalum (Slime mold).
                Glutathione reductase"
                /protein_id="AAM09354.2"
                /db_xref="GI:28828055"
                /translation="MSSTNHFTYLVLGAGSGGIASARRAAKHLNAKGNGDRIGIVEVT
                RPGGTCVNVGCVPKKVMWNTSFIKEMINAAPSYGFDFGGQQVKFNWPTIKKARDEYIK
                RLNGIYDSNLAKDNIVRINGYGRFSGPKEIQVNGANGEKYTADHILIAAGGRPTVPDV
                PGKELGITSDGFFELEDLPKSTLVVGAGYIAVELAGVLHSLGSETTMVIRQKQFLRTF
                DEMLHTTLLKQMTDDGVKFVTEASIKSLERDVDGKRIIATTNAGVKLPPVECVIWAIG
                RVPNTDDLGIDKAGIQLTEQSGFIKVDEFQNTNVPGVHAVGDICGNFLLTPVAIAAGR
                RLSERLFNGKSDLKFEYENVATVVFSHPPIGTVGLTEQEAITKYGTENIKCYNTSFIN
                MFYSVQVHKVRTSMKLVCLGKEEKVIGLHIIGDGCDEIIQGFAVAVKMGCTKWDLDNT
                CAIHPTSAEELVTMV"
     CDS        complement(join(18106. .18501,18638. .18994))
                /note="GeneID exon scores (in order of location ranges):
                56.32, 49.02 - GSCJ_ID dd_00404"
                /codon_start=1

Alignment Scores:
Pred. No.:              119          Length:        254733
Score:                  73.00        Matches:       10
Percent Similarity:     91.67%       Conservative:  1
Best Local Similarity:  83.33%       Mismatches:    1
Query Match:            86.90%       Indels:        0
DB:                     2            Gaps:          0

US-09-858-332G-15 (1-12) x AC117075 (1-254733)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||||||||   ||||||||||||||:::||||||
Db       1929 CACAACCACAAACACAACCACAATCACAGTCACAAT 1894

RESULT 285
BTA512824/c
LOCUS       BTA512824               190 bp    mRNA    linear   MAM 28-JUL-2004
DEFINITION  Bos taurus partial mRNA for immunoglobulin heavy chain variable
            region (IGHV gene).
ACCESSION   AJ512824
VERSION     AJ512824.1  GI:24412782
KEYWORDS    IGHV gene; immunoglobulin heavy chain; variable region.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1
  AUTHORS   Panelli,S., Damiani,G., Galli,C. and Sgaramella,V.
  TITLE     Rearranged genomes of bovine blood cells can allow the development
            of clones till late fetal stages; but rare unrearranged genomes
            have greater potential and lead to adulthood
  JOURNAL   Gene 334, 99-103 (2004)
   PUBMED   15256259
REFERENCE   2  (bases 1 to 190)
  AUTHORS   Panelli,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2002) Panelli S., Patol. Anim. Igiene e Sanita
            Pubbl. Vet., Fondazione PTP c/o Universita di Milano, Via Celoria
            10, Milano 20133, ITALY
FEATURES             Location/Qualifiers
     source          1. .190
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /cell_type="lymphocytes"
                     /rearranged
     gene            1. .190
                     /gene="IGHV"
     CDS             <1. .>190
                     /gene="IGHV"
                     /codon_start=2
                     /product="immunoglobulin heavy chain variable region"
                     /protein_id="CAD55149.1"
                     /db_xref="GI:24412783"
                     /translation="PGLKSRLSITKDNSKSQVFLSLSSLTPEDTATYYCASMVVMVMV
                     VMVVMVMVMVMVYVDAWGQ"
     V_region        <1. .>190
                     /gene="IGHV"
     V_segment       1. .97
                     /gene="IGHV"
     J_segment       98. .190
                     /gene="IGHV"
ORIGIN

Alignment Scores:
Pred. No.:              0.156        Length:        190
Score:                  72.00        Matches:       12
Percent Similarity:     75.00%       Conservative:  0
Best Local Similarity:  75.00%       Mismatches:    0
Query Match:            85.71%       Indels:        4
DB:                     4            Gaps:          1

US-09-858-332G-15 (1-12) x BTA512824 (1-190)

Qy          1 HisAsnHisAsnHisAsnHisAsn----------HisAsnHisAsn 12
              ||||||||||||||||||||||||          ||||||||||||
Db        166 CATAACCATAACCATAACCATAACAACCATAACCACCATAACCATAAC 119

RESULT 286
BTA512976/c
LOCUS       BTA512976               209 bp    DNA     linear   MAM 28-JUL-2004
DEFINITION  Bos taurus partial IGHV gene for immunoglobulin heavy chain
            variable region, exon 1, isolate adult 1, clone 2.
ACCESSION   AJ512976
VERSION     AJ512976.1  GI:24459870
KEYWORDS    IGHV gene; immunoglobulin heavy chain; variable region.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1
  AUTHORS   Panelli,S., Damiani,G., Galli,C. and Sgaramella,V.
  TITLE     Rearranged genomes of bovine blood cells can allow the development
            of clones till late fetal stages; but rare unrearranged genomes
            have greater potential and lead to adulthood
  JOURNAL   Gene 334, 99-103 (2004)
   PUBMED   15256259
REFERENCE   2  (bases 1 to 209)
  AUTHORS   Panelli,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT-2002) Panelli S., Patol. Anim. Igiene e Sanita
            Pubbl. Vet., Fondazione PTP c/o Universita di Milano, via Celoria
            10, Milano 20133, ITALY
FEATURES             Location/Qualifiers
     source          1. .209
                     /organism="Bos taurus"
                     /mol_type="genomic DNA"
                     /isolate="adult 1"
                     /db_xref="taxon:9913"
                     /clone="2"
                     /cell_type="lymphocytes"
                     /dev_stage="adult"
                     /rearranged
     gene            1. .209
                     /gene="IGHV"
     CDS             <1. .>209
                     /gene="IGHV"
                     /codon_start=3
                     /product="immunoglobulin heavy chain variable region"
                     /protein_id="CAD55894.1"
                     /db_xref="GI:24459871"
                     /translation="QAGNPGSASPRTTPRAKSLGSVSSVTTEDSATYYCTTVHQIRSY
                     GGYGYGGYGCYGYGYGYDYVDAWGQ"
     exon            <1. .209
                     /gene="IGHV"
                     /number=1
     V_region        <1. .>209
                     /gene="IGHV"
                     /product="immunoglobulin heavy chain variable region"
     V_segment       1. .108
                     /gene="IGHV"
     J_segment       109. .209
                     /gene="IGHV"
ORIGIN

Alignment Scores:
Pred. No.:              0.171          Length:         209
Score:                  72.00          Matches:        12
Percent Similarity:     75.00%         Conservative:   0
Best Local Similarity:  75.00%         Mismatches:     0
Query Match:            85.71%         Indels:         4
DB:                     4              Gaps:           1

US-09-858-332G-15 (1-12) x BTA512976 (1-209)

Qy          1 HisAsnHisAsnHisAsnHisAsn-----------HisAsnHisAsn 12
              |||||||||||||||||||||||           ||||||||||||
Db        186 CATAACCATAACCATAACCATAACAACCATAACCACCATAACCATAAC 139

RESULT 287
BTU49777/c
LOCUS       BTU49777                 360 bp    mRNA    linear   MAM 16-AUG-2001
DEFINITION  Bos taurus immunoglobulin rearranged heavy chain variable region
            mRNA, partial cds.
ACCESSION   U49777
VERSION     U49777.1  GI:1293599
KEYWORDS    .
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 360)
  AUTHORS   Berens,S.J., Wylie,D.E. and Lopez,O.J.
  TITLE     Use of a single VH family and long CDR3s in the variable region of
            cattle Ig heavy chains
  JOURNAL   Int. Immunol. 9 (1), 189-199 (1997)
   PUBMED   9043960
REFERENCE   2  (bases 1 to 360)
  AUTHORS   Berens,S.J., Wylie,D. and Lopez,O.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-1996) Osvaldo J. Lopez, Biological Sciences,
            University of Nebraska, 325 Manter Hall, Lincoln, NE 68588-0118,
            USA
FEATURES             Location/Qualifiers
     source          1. .360
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /cell_type="heterohybridoma"
                     /tissue_type="spleen"
                     /dev_stage="150 day fetus"
     CDS             <1. .>360
                     /codon_start=1
                     /product="immunoglobulin heavy chain variable region"
                     /protein_id="AAA98658.1"
                     /db_xref="GI:1293600"
                     /translation="LVKPSQTLSLTCTVSGFSLSSYGVGWVRQAPGKALECLGGISSG
                     GSTGYNPALKSRLSITKDNSKSQVSLSLSSVTTEDTATYYCAKYGGYGYGGYGCYGYG
                     YGYGIDAWGQGLLVTVSS"
ORIGIN

Alignment Scores:
Pred. No.:              0.289          Length:         360
Score:                  72.00          Matches:        12
Percent Similarity:     75.00%         Conservative:   0
Best Local Similarity:  75.00%         Mismatches:     0
Query Match:            85.71%         Indels:         4
DB:                     4              Gaps:           1

US-09-858-332G-15 (1-12) x BTU49777 (1-360)

Qy          1 HisAsnHisAsnHisAsnHisAsn-----------HisAsnHisAsn 12
              |||||||||||||||||||||||           ||||||||||||
Db        316 CATAACCATAACCATAACCATAACAACCATAACCACCATAACCATAAC 269

RESULT 288
DQ120961
LOCUS       DQ120961                 1239 bp    DNA     linear   VRT 02-AUG-2005
DEFINITION  Hucho taimen clone ZB-12 microsatellite sequence.
ACCESSION   DQ120961
VERSION     DQ120961.1  GI:71385014
KEYWORDS    .
SOURCE      Hucho taimen
  ORGANISM  Hucho taimen
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Protacanthopterygii; Salmoniformes; Salmonidae; Hucho.
REFERENCE   1  (bases 1 to 1239)
  AUTHORS   Sun,X., Tong,G., Lu,C., Kuang,Y. and Liang,L.
  TITLE     Microsatellite Enrichment by Magnetic Beads in taimen
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1239)
  AUTHORS   Sun,X., Tong,G., Lu,C., Kuang,Y. and Liang,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUL-2005) Biotech Breeding Lab, Heilongjiang River
            Fishery Research Institute of Chinese Academy of Fishery Science,
            No. 43 Songfa Road Daoli District, Harbin, Heilongjiang 150070,
            China
FEATURES             Location/Qualifiers
     source          1. .1239
                     /organism="Hucho taimen"
                     /mol_type="genomic DNA"
```

```
                     /db_xref="taxon:201498"
                     /clone="ZB-12"
     repeat_region   224. .747
                     /note="microsatellite"
                     /rpt_type=tandem
ORIGIN

Alignment Scores:
Pred. No.:                0.95          Length:        1239
Score:                    72.00         Matches:       10
Percent Similarity:       100.00%       Conservative:  2
Best Local Similarity:    83.33%        Mismatches:    0
Query Match:              85.71%        Indels:        0
DB:                       5             Gaps:          0

US-09-858-332G-15 (1-12) x DQ120961 (1-1239)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||::::::|||||||||||||||||||||||||||||
Db        642 CACCACAACAACCATAACCACAACCACAACCACAAC 677

RESULT 289
AR549509
LOCUS       AR549509               1953 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 4640 from patent US 6747137.
ACCESSION   AR549509
VERSION     AR549509.1  GI:53942684
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1953)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 4640 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .1953
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:                1.47          Length:        1953
Score:                    72.00         Matches:       10
Percent Similarity:       100.00%       Conservative:  2
Best Local Similarity:    83.33%        Mismatches:    0
Query Match:              85.71%        Indels:        0
DB:                       6             Gaps:          0

US-09-858-332G-15 (1-12) x AR549509 (1-1953)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              :::||||||||||||||||||||||||||||||:::
Db       1685 AACAATCACAATCACAATCACAATCACAATCACAGT 1720

RESULT 290
AX536944
LOCUS       AX536944               3042 bp    DNA     linear   PAT 22-NOV-2002
DEFINITION  Sequence 545 from Patent WO02064766.
ACCESSION   AX536944
VERSION     AX536944.1  GI:25263379
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1
  AUTHORS   Contreras,R.H., Eberhardt,I., Luyten,W.H. and Reekmans,R.J.
  TITLE     Bax-responsive genes for drug target identification in yeast and
            fungi
  JOURNAL   Patent: WO 02064766-A 545 22-AUG-2002;
            JANSSEN PHARMACEUTICA N.V. (BE)
FEATURES             Location/Qualifiers
     source          1. .3042
                     /organism="Candida albicans"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:5476"
ORIGIN

Alignment Scores:
Pred. No.:                2.26          Length:        3042
Score:                    72.00         Matches:       10
Percent Similarity:       100.00%       Conservative:  2
Best Local Similarity:    83.33%        Mismatches:    0
Query Match:              85.71%        Indels:        0
DB:                       6             Gaps:          0

US-09-858-332G-15 (1-12) x AX536944 (1-3042)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              :::||||||||||||||||||||||||||||||:::
Db       2690 AACAATCACAATCACAATCACAATCACAATCACAGT 2725

RESULT 291
AX345465/c
LOCUS       AX345465               5826 bp    DNA     linear   PAT 01-FEB-2002
DEFINITION  Sequence 536 from Patent WO0200928.
ACCESSION   AX345465
VERSION     AX345465.1  GI:18493351
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with the immune system
  JOURNAL   Patent: WO 0200928-A 536 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .5826
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

Alignment Scores:
Pred. No.:                4.22          Length:        5826
Score:                    72.00         Matches:       10
Percent Similarity:       90.91%        Conservative:  0
Best Local Similarity:    90.91%        Mismatches:    1
Query Match:              85.71%        Indels:        0
DB:                       6             Gaps:          0

US-09-858-332G-15 (1-12) x AX345465 (1-5826)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              |||||||||||||||   |||||||||||||||
Db        335 CATAACCACAACCATACCCATAACCATAACCAC 303

RESULT 292
AX344794/c
LOCUS       AX344794               6236 bp    DNA     linear   PAT 01-FEB-2002
DEFINITION  Sequence 219 from Patent WO0200927.
ACCESSION   AX344794
VERSION     AX344794.1  GI:18492680
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with development genes
  JOURNAL   Patent: WO 0200927-A 219 03-JAN-2002;
```

```
           Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .6236
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

Alignment Scores:
Pred. No.:              4.51          Length:         6236
Score:                  72.00         Matches:        10
Percent Similarity:     90.91%        Conservative:   0
Best Local Similarity:  90.91%        Mismatches:     1
Query Match:            85.71%        Indels:         0
DB:                     6             Gaps:           0

US-09-858-332G-15 (1-12) x AX344794 (1-6236)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              ||||||||||||||||   ||||||||||||||||
Db       4241 CATAACCACAACCATACCCATAACCATAACCAC 4209

RESULT 293
AX346532/c
LOCUS       AX346532               6236 bp    DNA     linear   PAT 01-FEB-2002
DEFINITION  Sequence 1603 from Patent WO0200928.
ACCESSION   AX346532
VERSION     AX346532.1  GI:18494418
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with the immune system
  JOURNAL   Patent: WO 0200928-A 1603 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .6236
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

Alignment Scores:
Pred. No.:              4.51          Length:         6236
Score:                  72.00         Matches:        10
Percent Similarity:     90.91%        Conservative:   0
Best Local Similarity:  90.91%        Mismatches:     1
Query Match:            85.71%        Indels:         0
DB:                     6             Gaps:           0

US-09-858-332G-15 (1-12) x AX346532 (1-6236)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              ||||||||||||||||   ||||||||||||||||
Db       4241 CATAACCACAACCATACCCATAACCATAACCAC 4209

RESULT 294
AC165067/c
LOCUS       AC165067               39156 bp    DNA     linear   PLN 26-JUL-2005
DEFINITION  Phakopsora pachyrhizi clone JGIAFNA-421A12, complete sequence.
ACCESSION   AC165067
VERSION     AC165067.2  GI:71143210
KEYWORDS    HTG.
SOURCE      Phakopsora pachyrhizi
  ORGANISM  Phakopsora pachyrhizi
            Eukaryota; Fungi; Basidiomycota; Urediniomycetes;
            Urediniomycetidae; Uredinales; Phakopsoraceae; Phakopsora.
REFERENCE   1  (bases 1 to 39156)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 39156)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-2005) Production Genomics Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA
            94598-1698, USA
REFERENCE   3  (bases 1 to 39156)
  AUTHORS   Stanford Human Genome Center.
  CONSRTM   DOE Joint Genome Institute
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2005) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jul 26, 2005 this sequence version replaced gi:68533351.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 100% of Sequence;
            Estimated Total Number of Errors is 0.
FEATURES             Location/Qualifiers
     source          1. .39156
                     /organism="Phakopsora pachyrhizi"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:170000"
                     /clone="JGIAFNA-421A12"
ORIGIN

Alignment Scores:
Pred. No.:              26.4          Length:         39156
Score:                  72.00         Matches:        10
Percent Similarity:     90.91%        Conservative:   0
Best Local Similarity:  90.91%        Mismatches:     1
Query Match:            85.71%        Indels:         0
DB:                     15            Gaps:           0

US-09-858-332G-15 (1-12) x AC165067 (1-39156)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              |||   ||||||||||||||||||||||||||||
Db      18992 CATCAACATAACCACAACCACAACCACAACCAC 18960

RESULT 295
AL845551/c
LOCUS       AL845551               40374 bp    DNA     linear   VRT 16-APR-2005
DEFINITION  Zebrafish DNA sequence from clone XX-119C3 Contains 3 novel genes
            for alpha-type globins and one novel gene for a beta-type globin,
            complete sequence.
ACCESSION   AL845551
VERSION     AL845551.2  GI:23380956
KEYWORDS    HTG; globin; hemoglobin.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 40374)
  AUTHORS   Matthews,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-DEC-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests:
            http://www.sanger.ac.uk/Projects/D_rerio/faqs.shtml#dataeight
COMMENT     On Sep 30, 2002 this sequence version replaced gi:22416327.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep XX-119C3 is from
            a Incyte Zebrafish AB BAC library VECTOR: pBeloBAC11
            Clone-derived Zebrafish pUC subclones occasionally display
            inconsistency over the length of mononucleotide A/T runs and
```

```
            conserved TA repeats. Where this is found the longest good quality
            representation will be submitted.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
            Any regions longer than 1kb tagged as misc-feature 'unsure' are
            part of a tandem repeat of more than 10kb in length where it has
            not been possible to anchor the base differences between repeat
            copies.  The region has been built up based on the repeat element
            to match the total size of repeat indicated by restriction digest,
            but repeat copies may not be in the correct order and the usual
            finishing criteria may not apply.
FEATURES             Location/Qualifiers
     source          1. .40374
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="XX-119C3"
                     /clone_lib="Incyte-BAC"
     misc_feature    1. .48
                     /note="Sequence from uni-directional dGTP big dye
                     terminator reads only."
     misc_feature    1
                     /note="Clone_left_end: XX-119C3"
     gene            complement(3757. .4755)
                     /locus_tag="bY119C3.1-001"
     mRNA            complement(join(3757. .4252,4335. .4542,4624. .4755))
                     /locus_tag="bY119C3.1-001"
                     /product="novel alpha-type globin"
                     /note="match: cDNAs: AB015448.1 AB015449.1 AB026052.1
                     AB080117.1
                     match: ESTs: AL720011.1 AL727420.1 AL721548.1 BQ783706.1
                     AL727315.1 AL727714.1 BE693186.1 AL720600.1 AI617394.1
                     AL719910.1 AL727764.1"
     CDS             complement(join(4124. .4252,4335. .4542,4624. .4718))
                     /locus_tag="bY119C3.1-001"
                     /standard_name="OTTDARP00000001843"
                     /note="match: proteins: Q8UW92 BAC20290 BAC20292 O13169
                     BAC20294 BAC20295"
                     /codon_start=1
                     /product="novel alpha-type globin"
                     /protein_id="CAD60790.1"
                     /db_xref="GI:27802786"
                     /db_xref="GOA:Q7ZT21"
                     /db_xref="InterPro:IPR000971"
                     /db_xref="InterPro:IPR002338"
                     /db_xref="InterPro:IPR002339"
                     /db_xref="UniProt/TrEMBL:Q7ZT21"
                     /db_xref="ZFIN:ZDB-GENE-980526-80"
                     /translation="MSLSAKDKAAVKTLWAKIAGKADDIGHDALSRMLIVYPQTKTYF
                     SHWKDLSPGSAPVRKHGKTVMGGVAEAVSKIDDLNAGLLNLSELHAFQLRVDPANFKI
                     LSHNILVVLATLFPADFTPEAHVAMDKFLSALALAMSEKYR"
     gene            join(7826. .7974,8053. .8816)
                     /locus_tag="bY119C3.2-002"
     gene            join(7826. .7974,8053. .8275,8370. .8816)
                     /locus_tag="bY119C3.2-001"
     mRNA            join(7826. .7974,8053. .8816)
                     /locus_tag="bY119C3.2-002"
                     /product="novel beta-type globin"
                     /note="match: ESTs: AL727267.1"
     mRNA            join(7826. .7974,8053. .8275,8370. .8816)
                     /locus_tag="bY119C3.2-001"
                     /product="novel beta-type globin"
                     /note="match: cDNAs: AF082662.1
                     match: ESTs: BM341906.1 AL909871.1 AI943217.1 AI965114.1
                     BM534785.1"
     CDS             join(7883. .7974,8053. .8275,8370. .8498)
                     /locus_tag="bY119C3.2-001"
                     /standard_name="OTTDARP00000001844"
                     /note="match: proteins: Q98852 O93548 BAC20296"
                     /codon_start=1
                     /product="novel beta-type globin"
                     /protein_id="CAD60792.1"
                     /db_xref="GI:27802788"
                     /db_xref="GOA:O93548"
                     /db_xref="InterPro:IPR000971"
                     /db_xref="InterPro:IPR002337"
                     /db_xref="UniProt/TrEMBL:O93548"
                     /db_xref="ZFIN:ZDB-GENE-030616-7"
                     /translation="MVVWTDFEKATIQDIFAKADYDVIGPQALARCLIVYPWTQRYFA
                     KFGNLYNAAAILGNPMVAAHGKTVLKGLELAVKNMDNIKATYADLSVLHSEKLHVDPD
                     NFRLLADCLTIVVAAQMGAGFTPEVQAAFQKFIAVAVSALGRQYH"
     CDS             join(7883. .7974,8053. .8284)
                     /locus_tag="bY119C3.2-002"
                     /standard_name="OTTDARP00000001845"
                     /codon_start=1
                     /product="novel beta-type globin"
                     /protein_id="CAD60791.1"
                     /db_xref="GI:27802787"
                     /db_xref="GOA:Q7ZYZ5"
                     /db_xref="InterPro:IPR000971"
                     /db_xref="InterPro:IPR002337"
                     /db_xref="UniProt/TrEMBL:Q7ZYZ5"
                     /db_xref="ZFIN:ZDB-GENE-030616-7"
                     /translation="MVVWTDFEKATIQDIFAKADYDVIGPQALARCLIVYPWTQRYFA
                     KFGNLYNAAAILGNPMVAAHGKTVLKGLELAVKNMDNIKATYADLSVLHSEKLHVDPD
                     NFRVS"
     gene            complement(14921. .15637)
                     /locus_tag="bY119C3.3-001"
     mRNA            complement(join(14921. .15137,15217. .15424,15506. .15637))
                     /locus_tag="bY119C3.3-001"
                     /product="novel alpha-type globin"
                     /note="match: cDNAs: AB015448.1 AB026052.1 AB080117.1
                     match: ESTs: AL721548.1 BQ783706.1 BQ618695.1 AI545160.1
                     AI617394.1 AL719910.1 AL721636.1 AL911459.1"
     CDS             complement(join(15009. .15137,15217. .15424,15506. .15600))
                     /locus_tag="bY119C3.3-001"
                     /standard_name="OTTDARP00000001846"
                     /note="match: proteins: Q90487 BAC20290 P83270 Q8UW95
                     O13169"
                     /codon_start=1
                     /product="novel alpha-type globin"
                     /protein_id="CAD60793.1"
                     /db_xref="GI:27802789"
                     /db_xref="GOA:Q7ZT21"
                     /db_xref="InterPro:IPR000971"
                     /db_xref="InterPro:IPR002338"
                     /db_xref="InterPro:IPR002339"
                     /db_xref="UniProt/TrEMBL:Q7ZT21"
                     /db_xref="ZFIN:ZDB-GENE-980526-80"
                     /translation="MSLSAKDKAAVKTLWAKIAGKADDIGHDALSRMLIVYPQTKTYF
                     SHWKDLSPGSAPVRKHGKTVMGGVAEAVSKIDDLNAGLLNLSELHAFQLRVDPANFKI
                     LSHNILVVLATLFPADFTPEAHVAMDKFLSALALAMSEKYR"
     gene            complement(23127. .24272)
                     /locus_tag="bY119C3.4-001"
     mRNA            complement(join(23127. .23351,23715. .23922,24142. .24272))
                     /locus_tag="bY119C3.4-001"
                     /product="novel alpha-type globin"
                     /note="match: cDNAs: AB015448.1 AB034640.1 AB015449.1
                     match: ESTs: AI477488.1 AL912768.1 AL912769.1 AI353083.1
                     AI353675.1 AI354176.1"
     CDS             complement(join(23223. .23351,23715. .23922,24142. .24236))
                     /locus_tag="bY119C3.4-001"
                     /standard_name="OTTDARP00000001847"
                     /note="match: proteins: BAC20301 Q9YGF7 BAC20294 O13137"
                     /codon_start=1
```

```
                     /product="novel alpha-type globin"
                     /protein_id="CAD60794.1"
                     /db_xref="GI:27802790"
                     /db_xref="GOA:Q7ZYZ4"
                     /db_xref="InterPro:IPR000971"
                     /db_xref="InterPro:IPR002338"
                     /db_xref="UniProt/TrEMBL:Q7ZYZ4"
                     /db_xref="ZFIN:ZDB-GENE-990706-3"
                     /translation="MSLSAKDKANVKAFFDKVAPKAEEIGRETLSRTLFVYPQTKTYF
                     SHWADLSPNSPQVKKHGTTVINGVLTAVGLMDDLKGGLLTLSELHAFMLRVDPANFKI
                     INHNLLVSLAMMFPDDFTPEVHVSVDKFLAQVSLALSEKYR"
     misc_feature    40374
                     /note="Clone_right_end: XX-119C3"
ORIGIN

Alignment Scores:
Pred. No.:                27.2          Length:          40374
Score:                    72.00         Matches:         10
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    90.91%        Mismatches:      0
Query Match:              85.71%        Indels:          0
DB:                       5             Gaps:            0

US-09-858-332G-15 (1-12) x AL845551 (1-40374)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              ||||||||||||||||||||||:::||||||
Db      10229 CATAATCATAATCATAATCATAATTATAATCAT 10197

RESULT 296
AC021929/c
LOCUS       AC021929               81767 bp    DNA     linear   HTG 24-AUG-2002
DEFINITION  Homo sapiens chromosome 11 clone RP11-128A23 map 11, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC021929
VERSION     AC021929.2  GI:9141461
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 81767)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-128A23
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 81767)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 81767)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 13, 2000 this sequence version replaced gi:6730844.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L4942
                Center clone name: 128_A_23
            --------------.
            * NOTE: This record contains 85 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      867: contig of 867 bp in length
            *      868      967: gap of 100 bp
            *      968     1849: contig of 882 bp in length
            *     1850     1949: gap of 100 bp
            *     1950     2828: contig of 879 bp in length
            *     2829     2928: gap of 100 bp
            *     2929     3777: contig of 849 bp in length
            *     3778     3877: gap of 100 bp
            *     3878     4722: contig of 845 bp in length
            *     4723     4822: gap of 100 bp
            *     4823     5671: contig of 849 bp in length
            *     5672     5771: gap of 100 bp
            *     5772     6633: contig of 862 bp in length
            *     6634     6733: gap of 100 bp
            *     6734     7585: contig of 852 bp in length
            *     7586     7685: gap of 100 bp
            *     7686     8538: contig of 853 bp in length
            *     8539     8638: gap of 100 bp
            *     8639     9495: contig of 857 bp in length
            *     9496     9595: gap of 100 bp
            *     9596    10473: contig of 878 bp in length
            *    10474    10573: gap of 100 bp
            *    10574    11414: contig of 841 bp in length
            *    11415    11514: gap of 100 bp
            *    11515    12370: contig of 856 bp in length
            *    12371    12470: gap of 100 bp
            *    12471    13293: contig of 823 bp in length
            *    13294    13393: gap of 100 bp
            *    13394    14266: contig of 873 bp in length
            *    14267    14366: gap of 100 bp
```

```
*    14367    15220: contig of 854 bp in length
*    15221    15320: gap of 100 bp
*    15321    16188: contig of 868 bp in length
*    16189    16288: gap of 100 bp
*    16289    17155: contig of 867 bp in length
*    17156    17255: gap of 100 bp
*    17256    18107: contig of 852 bp in length
*    18108    18207: gap of 100 bp
*    18208    19067: contig of 860 bp in length
*    19068    19167: gap of 100 bp
*    19168    20011: contig of 844 bp in length
*    20012    20111: gap of 100 bp
*    20112    20989: contig of 878 bp in length
*    20990    21089: gap of 100 bp
*    21090    21968: contig of 879 bp in length
*    21969    22068: gap of 100 bp
*    22069    22905: contig of 837 bp in length
*    22906    23005: gap of 100 bp
*    23006    23870: contig of 865 bp in length
*    23871    23970: gap of 100 bp
*    23971    24829: contig of 859 bp in length
*    24830    24929: gap of 100 bp
*    24930    25780: contig of 851 bp in length
*    25781    25880: gap of 100 bp
*    25881    26755: contig of 875 bp in length
*    26756    26855: gap of 100 bp
*    26856    27705: contig of 850 bp in length
*    27706    27805: gap of 100 bp
*    27806    28665: contig of 860 bp in length
*    28666    28765: gap of 100 bp
*    28766    29612: contig of 847 bp in length
*    29613    29712: gap of 100 bp
*    29713    30599: contig of 887 bp in length
*    30600    30699: gap of 100 bp
*    30700    31560: contig of 861 bp in length
*    31561    31660: gap of 100 bp
*    31661    32530: contig of 870 bp in length
*    32531    32630: gap of 100 bp
*    32631    33489: contig of 859 bp in length
*    33490    33589: gap of 100 bp
*    33590    34458: contig of 869 bp in length
*    34459    34558: gap of 100 bp
*    34559    35421: contig of 863 bp in length
*    35422    35521: gap of 100 bp
*    35522    36382: contig of 861 bp in length
*    36383    36482: gap of 100 bp
*    36483    37350: contig of 868 bp in length
*    37351    37450: gap of 100 bp
*    37451    38308: contig of 858 bp in length
*    38309    38408: gap of 100 bp
*    38409    39266: contig of 858 bp in length
*    39267    39366: gap of 100 bp
*    39367    40248: contig of 882 bp in length
*    40249    40348: gap of 100 bp
*    40349    41207: contig of 859 bp in length
*    41208    41307: gap of 100 bp
*    41308    42195: contig of 888 bp in length
*    42196    42295: gap of 100 bp
*    42296    43168: contig of 873 bp in length
*    43169    43268: gap of 100 bp
*    43269    44151: contig of 883 bp in length
*    44152    44251: gap of 100 bp
*    44252    45106: contig of 855 bp in length
*    45107    45206: gap of 100 bp
*    45207    46044: contig of 838 bp in length
*    46045    46144: gap of 100 bp
*    46145    47011: contig of 867 bp in length
*    47012    47111: gap of 100 bp
*    47112    47971: contig of 860 bp in length
*    47972    48071: gap of 100 bp
*    48072    48936: contig of 865 bp in length
*    48937    49036: gap of 100 bp
*    49037    49903: contig of 867 bp in length
*    49904    50003: gap of 100 bp
*    50004    50824: contig of 821 bp in length
*    50825    50924: gap of 100 bp
*    50925    51792: contig of 868 bp in length
*    51793    51892: gap of 100 bp
*    51893    52777: contig of 885 bp in length
*    52778    52877: gap of 100 bp
*    52878    53722: contig of 845 bp in length
*    53723    53822: gap of 100 bp
*    53823    54669: contig of 847 bp in length
*    54670    54769: gap of 100 bp
*    54770    55628: contig of 859 bp in length

Alignment Scores:
Pred. No.:              53.7          Length:         81767
Score:                  72.00         Matches:        10
Percent Similarity:     100.00%       Conservative:   2
Best Local Similarity:  83.33%        Mismatches:     0
Query Match:            85.71%        Indels:         0
DB:                     14            Gaps:           0

US-09-858-332G-15 (1-12) x AC021929 (1-81767)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              ||||||||||||:::||||||||||||:::||||||
Db      13091 CATAATCATAATAATAATCATAATCATCATCATAAT 13056

RESULT 297
CR380953_01/c
WPCOMMENT
  Sequence split into 10 fragments  LOCUS CR380953 Accession CR380953
     Fragment Name      Begin     End
     CR380953_00            1    110000
     CR380953_01       100001    210000
     CR380953_02       200001    310000
     CR380953_03       300001    410000
     CR380953_04       400001    510000
     CR380953_05       500001    610000
     CR380953_06       600001    710000
     CR380953_07       700001    810000
     CR380953_08       800001    910000
     CR380953_09       900001    992211
Continuation (2 of 10) of CR380953 from base 100001 (CR380953 Candida glabrata strain CBS

Alignment Scores:
Pred. No.:              71.5          Length:         110000
Score:                  72.00         Matches:        10
Percent Similarity:     100.00%       Conservative:   2
Best Local Similarity:  83.33%        Mismatches:     0
Query Match:            85.71%        Indels:         0
DB:                     15            Gaps:           0

US-09-858-332G-15 (1-12) x CR380953_01 (1-110000)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              :::||||||||||||||||||||||||||:::|||
Db      66957 TACAACCACAACCATAACCACAACCACAACTACAAC 66922

RESULT 298
CR382129_00/c
WPCOMMENT
  Sequence split into 33 fragments  LOCUS CR382129 Accession CR382129
     Fragment Name      Begin     End
     CR382129_00            1    110000
     CR382129_01       100001    210000
     CR382129_02       200001    310000
     CR382129_03       300001    410000
     CR382129_04       400001    510000
     CR382129_05       500001    610000
     CR382129_06       600001    710000
     CR382129_07       700001    810000
     CR382129_08       800001    910000
     CR382129_09       900001   1010000
```

```
     CR382129_10       1000001   1110000
     CR382129_11       1100001   1210000
     CR382129_12       1200001   1310000
     CR382129_13       1300001   1410000
     CR382129_14       1400001   1510000
     CR382129_15       1500001   1610000
     CR382129_16       1600001   1710000
     CR382129_17       1700001   1810000
     CR382129_18       1800001   1910000
     CR382129_19       1900001   2010000
     CR382129_20       2000001   2110000
     CR382129_21       2100001   2210000
     CR382129_22       2200001   2310000
     CR382129_23       2300001   2410000
     CR382129_24       2400001   2510000
     CR382129_25       2500001   2610000
     CR382129_26       2600001   2710000
     CR382129_27       2700001   2810000
     CR382129_28       2800001   2910000
     CR382129_29       2900001   3010000
     CR382129_30       3000001   3110000
     CR382129_31       3100001   3210000
     CR382129_32       3200001   3272609
LOCUS       CR382129              3272609 bp    DNA     linear   PLN 17-APR-2005
DEFINITION  Yarrowia lipolytica chromosome C of strain CLIB99 of Yarrowia
            lipolytica.
ACCESSION   CR382129
VERSION     CR382129.1  GI:49647136
KEYWORDS    genomic DNA.
SOURCE      Yarrowia lipolytica CLIB99
  ORGANISM  Yarrowia lipolytica CLIB99
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Dipodascaceae; Yarrowia.
REFERENCE   1  (bases 1 to 3272609)
  AUTHORS   Dujon,B., Sherman,D., Fischer,G., Durrens,P., Casaregola,S.,
            Lafontaine,I., De Montigny,J., Marck,C., Neuveglise,C., Talla,E.,
            Goffard,N., Frangeul,L., Aigle,M., Anthouard,V., Babour,A.,
            Barbe,V., Barnay,S., Blanchin,S., Beckerich,J.M., Beyne,E.,
            Bleykasten,C., Boisrame,A., Boyer,J., Cattolico,L.,
            Confanioleri,F., De Daruvar,A., Despons,L., Fabre,E., Fairhead,C.,
            Ferry-Dumazet,H., Groppi,A., Hantraye,F., Hennequin,C.,
            Jauniaux,N., Joyet,P., Kachouri,R., Kerrest,A., Koszul,R.,
            Lemaire,M., Lesur,I., Ma,L., Muller,H., Nicaud,J.M., Nikolski,M.,
            Oztas,S., Ozier-Kalogeropoulos,O., Pellenz,S., Potier,S.,
            Richard,G.F., Straub,M.L., Suleau,A., Swennen,D., Tekaia,F.,
            Wesolowski-Louvel,M., Westhof,E., Wirth,B., Zeniou-Meyer,M.,
            Zivanovic,I., Bolotin-Fukuhara,M., Thierry,A., Bouchier,C.,
            Caudron,B., Scarpelli,C., Gaillardin,C., Weissenbach,J., Wincker,P.
            and Souciet,J.L.
  CONSRTM   Genolevures
  TITLE     Genome evolution in yeasts
  JOURNAL   Nature 430 (6995), 35-44 (2004)
   PUBMED   15229592
REFERENCE   2  (bases 1 to 3272609)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUL-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex -FRANCE (E-mail : seqref@genoscope.cns.fr -
            Web : www.genoscope.cns.fr)
COMMENT     This sequence is unfinished. Sequencing was performed at Genoscope
            and annotations were obtained by the genolevures Consortium.
FEATURES             Location/Qualifiers
     source          1. .3272609
                     /organism="Yarrowia lipolytica CLIB99"
                     /mol_type="genomic DNA"
                     /strain="CLIB99"
                     /db_xref="taxon:284591"
                     /chromosome="C"
                     /note="Genoscope sequence ID : YALI0CCHR"
     gene            11541. .12740
                     /locus_tag="YALI0C00110g"
     CDS             11541. .12740
                     /locus_tag="YALI0C00110g"
                     /note="unnamed protein product; weakly similar to
                     tr|Q9P731 Neurospora crassa Hypothetical 48.8 kDa protein,
                     hypothetical start"
                     /codon_start=1
                     /protein_id="CAG81565.1"
                     /db_xref="GI:49647137"
                     /db_xref="UniProt/TrEMBL:Q6CDJ2"
                     /translation="MALFINLLAVFVSLIIGGTIYLDRSFSELTNRVSWNDRHVYNKA
                     AEYNAASCKVSKAIPEACEDSIVDDSTGLVYFSCGETENRTKWFPGCNWWDKEGMTTN
                     SFYVYDPETRDFEKMNLDYHGEFVSHGLSLLKDPNNPDTRYIFSVNHKSTGSIISIFS
                     FQAGTNDVSHVGDIKDSRIYTPNSVTAYVEDDGSIAIMYTNDHLFRKGFMRHVEDSYG
                     PFMWGHVGNCNFKIEDGDIKGLSCSRVAVENSYPNGITHIPDTKEFLVADTRLGTLSS
                     FKWDVDTKSLNLQHSTPFGAALDNVHVIAGTKDVIVAGFPDEKQVFYKFRNMHDKSIK
                     VESLVFRANSRDNYTVPHVLQKTDGLHGLQVVTNYNYLPKHDRLLGSSVCSEGILICD
                     NIYGIIM"
     gene            complement(12761. .>14200)
                     /locus_tag="YALI0C00132g"
     CDS             complement(12761. .>14200)
                     /locus_tag="YALI0C00132g"
                     /note="unnamed protein product; similar to DEHA0D11396g
                     Debaryomyces hansenii IPF 1233.1, no start"
                     /codon_start=1
                     /protein_id="CAG81566.1"
                     /db_xref="GI:49647138"
                     /db_xref="GOA:Q6CDJ1"
                     /db_xref="InterPro:IPR011701"
                     /db_xref="UniProt/TrEMBL:Q6CDJ1"
                     /translation="TGVDSPRREETCAKLDWYVTTWACIMFVALQVDRGNLGQALTDN
                     LLDDLGMTTNDYNKGNTIFLCSFLAAEIPSQLISKKIGPDRWVPAQMVLWSVVAASQG
                     ALTGKGSFYATRCLLGLLEGGFIPDLVLWLSYFYTSKELPFRLSLFWGSQTLTTILTS
                     LMAFGLLRIHTSSIPHGWRWLFIIEGIMTLLIGLASFFKMPASPSQTKTWFRKNGWFT
                     EREEKIVVNRVLRDDPYKGDLHNREGLSVKRIWNAFADYRLWPIYAAGMFSQIATNAV
                     SFYMTLNLRGLGFSTFNTNLLTIPANVLHIIMMISITLSSEFFNERAFHCMAHPIWML
                     PGLMVLRFWKGALVDKWGTWAVMTIILGQPQNHAILVSWTSRNSNTVSQRTMGQSLYN
                     MFVQVGAIVGANIYRADDKPKYHRGNQQLIILYFILIVILAGTKLFYYVINKRRSEKW
                     NAMTKDEQQHYRNTNTDSGSKRLDFRLAL"
     exon            complement(12761. .14200)
                     /locus_tag="YALI0C00132g"
                     /number=2
     gene            complement(14047. .>14397)
                     /locus_tag="YALI0C00143g"
     CDS             complement(14047. .>14397)
                     /locus_tag="YALI0C00143g"
                     /note="unnamed protein product; no similarity"
                     /codon_start=1
                     /protein_id="CAG81567.1"
                     /db_xref="GI:49647139"
                     /db_xref="InterPro:IPR001005"
                     /db_xref="UniProt/TrEMBL:Q6CDJ0"
                     /translation="LAMADKKYDELSISVDSISDADTAAVSIIDGSGTGGNIFSNPEV
                     AKHFAKLYEDAQYENRHLFDPELEWTPQEEKKLVRSWTGMSPHGPVSCLWLSRLTEAT
                     LAKLLPITCLMIWE"
     exon            complement(14047. .14397)
                     /locus_tag="YALI0C00143g"
                     /number=2
     gene            complement(15006. .16343)
                     /locus_tag="YALI0C00165g"
     CDS             complement(15006. .16343)
                     /locus_tag="YALI0C00165g"
                     /note="unnamed protein product; similar to sp|Q9P371
                     Schizosaccharomyces pombe, hypothetical start"
                     /codon_start=1
                     /protein_id="CAG81568.1"
                     /db_xref="GI:49647140"
                     /db_xref="GOA:Q6CDI9"
                     /db_xref="InterPro:IPR000555"
                     /db_xref="InterPro:IPR003639"
                     /db_xref="UniProt/TrEMBL:Q6CDI9"
                     /translation="MLSIAEIAAKGLDYPYTNNTSFRVWIHVVGTVLREAEIYRDDGD
                     FEKAYLLYTRFADLLLNKMPGHPGLKANRALFKKMTGKMPGVLQEMQQLKKALEVQIT
                     EEKKAEEERQRKLRALQQASDSGRSGGSRGSGSTSRIDDLRSQLGQGGPDHVNLEDYS
                     FGRYDTRQVSEPMYEHGNWQQQAHHQHDTFDYPDLNRELSSVLSPYVYESGPLTLSPT
                     AGQIRLADSPPAIPAKPAALKASYEKPAAPAPPTTSSSSVEHTSRYTTEGGLALRSLF
```

```
                     IPAELEATFLKVAHANTVKNLETCGILCGKLSRNAFFVTHLMIPPQESTSDTCQTTNE
                     ELLFEQIDENDLFVLGWIHTHPTQTCFMSSVDLHTQNSYQIMLPEAVALVCAPQHDPN
                     FGIFRLSDPPGVDIIKNCNRGGFHPHTEDNIYNNARHAMVKNGLPFKMKDLRS"
     gene            complement(18351. .20855)
                     /locus_tag="YALI0C00187g"
     CDS             complement(18351. .20855)
                     /locus_tag="YALI0C00187g"
                     /note="unnamed protein product; weakly similar to
                     tr|Q8X227 Cryptococcus neoformans O-acetyltransferase or
                     tr|Q8X226 Cryptococcus neoformans O-acetyltransferase and
                     tr|Q96PB1 Homo sapiens C7ORF12 or tr|Q8WZ77 Homo sapiens
                     O-acetyltransferase, hypothetical start"
                     /codon_start=1
                     /protein_id="CAG81569.1"
                     /db_xref="GI:49647141"
                     /db_xref="UniProt/TrEMBL:Q6CDI8"
                     /translation="MMNNILGWFALFAFCGTVIRLLAPGDGYHCESTLESSGMWLQPA
                     VGRYQPPRNWQPKQCMLRTFDRASTAQCLTNETSLFIGDSHAREMFWAVVKDLDPNYD
                     TTNVEKHSNIEVAIDGVSLKFFWDPYFNNDGMFEMSKIANDQEQKERIKMVHISTGMW
                     YIHNKEDKDAGFKEWKQALRRLNEIVAAQDDAGRRALEFITVSPMLLPDYVTLTELKD
                     EAHLKGLDAQYLSRMNTYLDHNFNKNPYRKTRISVPFSFVDMRNKGGRAAFKRDGIHL
                     NTDVYKTMSDILRNVRCNQELLDHKYDITNGPVYPFSDTCCIHYKQPRTMGQVLLLGV
                     ICAILPIIYLGQRYTRSDYYPRFVRYVSWPVPSIQDETLLTACVVIAASLAYSFFCDR
                     TQFFGKSSKQFEASEFWVLILLFLVATGYSFEPQNDNSFLNRHQTEEWKGWMQIIILI
                     YHITGASKILPIYKFVRVLVAAYLFMTGFGHATFFIKRGDFSLKRATSVLFRMNFLSI
                     LLAYVMDTDYLFYYFAPLVSFWFCVVWITFRVLPSWNSIDTSVGPVLAKISASAVILN
                     ILVRFQLPFQIVFAVLKYLFNIQWNLREWRFRLILDIYIVYIGMFAAVATLRYKQGNF
                     PLKNLVTNRLVLGVVALITFIVYIAVAASFGIKQNYNQAHPYISWMPIVAPVILRNIT
                     VRFRSHYSTWMSFIGTCSLETFTLQFHFFLAADTKGRLYILNTPTSSGVLGMLQKYLN
                     FGLVTVIFIIVSHAVAEQSNALTKLFVGNQDQQLPEHVTAQSQPQGEQIDMGDSEPPV
                     STRRPPGRVVAFLKTHTLQVKIVVAFAFMWVFNMLP"
     gene            22948. .25131
                     /locus_tag="YALI0C00209g"
     CDS             22948. .25131
                     /locus_tag="YALI0C00209g"
                     /note="unnamed protein product; similar to sp|P32784
                     Saccharomyces cerevisiae YBL011w SCT1 suppresses a
                     choline-transport mutant, hypothetical start"
                     /codon_start=1
                     /protein_id="CAG81570.1"
                     /db_xref="GI:49647142"

Alignment Scores:
Pred. No.:              71.5          Length:        110000
Score:                  72.00         Matches:       10
Percent Similarity:     90.91%        Conservative:  0
Best Local Similarity:  90.91%        Mismatches:    1
Query Match:            85.71%        Indels:        0
DB:                     15            Gaps:          0

US-09-858-332G-15 (1-12) x CR382129_00 (1-110000)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              |||   ||||||||||||||||||||||||||
Db      86774 CACCAACACAACCACAACCACAACCACAACCAC 86742

RESULT 299
AC010468/c
LOCUS       AC010468              122102 bp    DNA     linear   PRI 01-MAY-2001
DEFINITION  Homo sapiens chromosome 5 clone CTD-2290C20, complete sequence.
ACCESSION   AC010468
VERSION     AC010468.5  GI:13899402
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 122102)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 122102)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 122102)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAY-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On May 1, 2001 this sequence version replaced gi:13676998.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.9% of Sequence;
            Estimated Total Number of Errors is 0.
FEATURES             Location/Qualifiers
     source          1. .122102
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-2290C20"
ORIGIN

Alignment Scores:
Pred. No.:              79.1          Length:        122102
Score:                  72.00         Matches:       10
Percent Similarity:     90.91%        Conservative:  0
Best Local Similarity:  90.91%        Mismatches:    1
Query Match:            85.71%        Indels:        0
DB:                     8             Gaps:          0

US-09-858-332G-15 (1-12) x AC010468 (1-122102)

Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHis 11
              ||||||||||||||||||||||||||   |||
Db      68394 CACAATCATAACCACAATCACAATCATACCCAT 68362

RESULT 300
CR855273
LOCUS       CR855273              161081 bp    DNA     linear   VRT 23-FEB-2005
DEFINITION  Zebrafish DNA sequence from clone DKEYP-34E1 in linkage group 24,
            complete sequence.
ACCESSION   CR855273
VERSION     CR855273.7  GI:59889991
KEYWORDS    HTG.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 161081)
  AUTHORS   Heath,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 16, 2005 this sequence version replaced gi:58865147.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
```

chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.
Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhirong Bao and Sean Eddy, submitted), and those beginning 'drr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml DKEYP-34E1 is from a Zebrafish BAC library
VECTOR: pIndigoBAC-5.

```
FEATURES             Location/Qualifiers
     source          1. .161081
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEYP-34E1"
                     /clone_lib="DanioKeypilot"
ORIGIN
```

Alignment Scores:

Pred. No.:	103	Length:	161081
Score:	72.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	2
Best Local Similarity:	83.33%	Mismatches:	0
Query Match:	85.71%	Indels:	0
DB:	5	Gaps:	0

US-09-858-332G-15 (1-12) x CR855273 (1-161081)

```
Qy          1 HisAsnHisAsnHisAsnHisAsnHisAsnHisAsn 12
              |||:::|||||||||||||||||||||||:::|||
Db      69705 CATCATCATAATCATAATCATAACCATAATAATAAT 69740
```

Search completed: November 28, 2005, 12:31:23
Job time : 5501 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 28, 2005, 10:10:26 ; Search time 3043 Seconds
(without alignments)
112.080 Million cell updates/sec

Title: US-09-858-332G-28
Perfect score: 42
Sequence: 1 HNHNHN 6

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 1000 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09858332/runat_23112005_103530_19835/app_query.fasta_1.199
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=1000
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=600 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09858332_@CGN_1_1_4939_@runat_23112005_103530_19835 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	42	100.0	21	6	AR579290	AR579290 Sequence
c	2	42	100.0	21	6	AX241197	AX241197 Sequence
c	3	42	100.0	21	6	AX487745	AX487745 Sequence
	4	42	100.0	42	6	AX594160	AX594160 Sequence
c	5	42	100.0	57	6	CQ754680	CQ754680 Sequence
c	6	42	100.0	57	6	AX813672	AX813672 Sequence
	7	42	100.0	65	6	AX483190	AX483190 Sequence
	8	42	100.0	65	6	AX485536	AX485536 Sequence
	9	42	100.0	100	6	AX594158	AX594158 Sequence
c	10	42	100.0	115	4	AY559840	AY559840 Bos tauru
c	11	42	100.0	115	6	AR568829	AR568829 Sequence
	12	42	100.0	122	4	AJ811567	AJ811567 Lama glam
c	13	42	100.0	139	6	AR568831	AR568831 Sequence
c	14	42	100.0	180	15	CNS01B16	AL113650 Botrytis
c	15	42	100.0	190	4	BTA512824	AJ512824 Bos tauru
	16	42	100.0	192	6	AR553849	AR553849 Sequence
c	17	42	100.0	205	4	AY559839	AY559839 Bos tauru
c	18	42	100.0	205	6	AR568828	AR568828 Sequence
c	19	42	100.0	209	4	BTA512976	AJ512976 Bos tauru
c	20	42	100.0	211	4	BTA512977	AJ512977 Bos tauru
c	21	42	100.0	244	15	AY095308	AY095308 Amaranthu
c	22	42	100.0	263	15	AY095306	AY095306 Amaranthu
c	23	42	100.0	267	15	AY095307	AY095307 Amaranthu
c	24	42	100.0	281	15	SL12421	Y12421 S.latifolia
c	25	42	100.0	292	6	AR014656	AR014656 Sequence
c	26	42	100.0	292	6	I18909	I18909 Sequence 7
c	27	42	100.0	311	4	BTU49780	U49780 Bos taurus
c	28	42	100.0	317	2	AJ748041	AJ748041 Trypanoso
	29	42	100.0	339	6	AR552556	AR552556 Sequence
c	30	42	100.0	346	4	AF058630	AF058630 Oryctolag
c	31	42	100.0	346	4	AF434459	AF434459 Oryctolag
c	32	42	100.0	349	4	AF434364	AF434364 Oryctolag
c	33	42	100.0	351	4	AF097208	AF097208 Bos tauru
	34	42	100.0	355	15	AY225843	AY225843 Amicia gl
c	35	42	100.0	360	4	BTU49777	U49777 Bos taurus
c	36	42	100.0	364	4	AF058641	AF058641 Oryctolag
c	37	42	100.0	366	15	AJ836260	AJ836260 Arabidops
c	38	42	100.0	372	9	MMU40980	U40980 Mus musculu
c	39	42	100.0	381	4	AF097206	AF097206 Bos tauru
c	40	42	100.0	387	4	AF097207	AF097207 Bos tauru
c	41	42	100.0	390	4	AF097205	AF097205 Bos tauru
c	42	42	100.0	390	4	AF097209	AF097209 Bos tauru
	43	42	100.0	397	2	PFA430086	AJ430086 Plasmodiu
c	44	42	100.0	408	4	BTU49781	U49781 Bos taurus
	45	42	100.0	409	2	PFA430087	AJ430087 Plasmodiu
	46	42	100.0	425	10	G46669	G46669 Z14484_1 Ze
	47	42	100.0	427	2	PFA430088	AJ430088 Plasmodiu
	48	42	100.0	431	10	BV211484	BV211484 BARC1076
c	49	42	100.0	432	6	AR547090	AR547090 Sequence
c	50	42	100.0	441	15	AY503292	AY503292 Sorghum b
c	51	42	100.0	447	10	G47457	G47457 Z24495_1 Ze
c	52	42	100.0	450	15	AY503290	AY503290 Sorghum b
c	53	42	100.0	450	15	AY503291	AY503291 Sorghum b
c	54	42	100.0	450	15	AY503293	AY503293 Sorghum b
c	55	42	100.0	450	15	AY503294	AY503294 Sorghum b
c	56	42	100.0	450	15	AY503295	AY503295 Sorghum b
c	57	42	100.0	450	15	AY503296	AY503296 Sorghum a
c	58	42	100.0	450	15	AY503297	AY503297 Sorghum b
c	59	42	100.0	450	15	AY503298	AY503298 Sorghum b
c	60	42	100.0	450	15	AY503299	AY503299 Sorghum b
c	61	42	100.0	450	15	AY503302	AY503302 Sorghum b
c	62	42	100.0	450	15	AY503304	AY503304 Sorghum b
c	63	42	100.0	450	15	AY503305	AY503305 Sorghum b
c	64	42	100.0	450	15	AY503306	AY503306 Sorghum b
c	65	42	100.0	450	15	AY503308	AY503308 Sorghum a
c	66	42	100.0	450	15	AY503309	AY503309 Sorghum a
c	67	42	100.0	450	15	AY503310	AY503310 Sorghum b
c	68	42	100.0	450	15	AY503311	AY503311 Sorghum b
c	69	42	100.0	450	15	AY503312	AY503312 Sorghum b
c	70	42	100.0	450	15	AY503313	AY503313 Sorghum b
c	71	42	100.0	450	15	AY503314	AY503314 Sorghum b
c	72	42	100.0	450	15	AY503315	AY503315 Sorghum b
c	73	42	100.0	451	15	AY503300	AY503300 Sorghum b
c	74	42	100.0	451	15	AY503303	AY503303 Sorghum b
c	75	42	100.0	451	15	AY503307	AY503307 Sorghum b
	76	42	100.0	456	6	AR508437	AR508437 Sequence

c	77	42	100.0	456	15	AY503301 AY503301 Sorghum b
	78	42	100.0	457	6	AR500762 AR500762 Sequence
	79	42	100.0	457	6	AR516044 AR516044 Sequence
c	80	42	100.0	458	10	BV331153 BV331153 S241P619R
	81	42	100.0	465	2	DVNEURS1 U12591 Drosophila
	82	42	100.0	467	10	BV004681 BV004681 S208P6706
c	83	42	100.0	477	10	BV213506 BV213506 S233P636F
c	84	42	100.0	478	4	BTU49779 U49779 Bos taurus
	85	42	100.0	478	10	BV281117 BV281117 S232P6238
c	86	42	100.0	480	4	BTU50559 U50559 Bos taurus
c	87	42	100.0	480	15	AY768786 AY768786 Passiflor
c	88	42	100.0	489	4	AF097203 AF097203 Bos tauru
c	89	42	100.0	492	4	AF097204 AF097204 Bos tauru
	90	42	100.0	519	9	AF187065 AF187065 Rattus no
	91	42	100.0	528	6	AR554223 AR554223 Sequence
	92	42	100.0	530	15	AY677214 AY677214 Campyloca
	93	42	100.0	536	15	AY677216 AY677216 Campyloca
	94	42	100.0	536	15	AY677217 AY677217 Campyloca
	95	42	100.0	536	15	AY677218 AY677218 Campyloca
	96	42	100.0	536	15	AY677219 AY677219 Campyloca
	97	42	100.0	536	15	AY677220 AY677220 Campyloca
	98	42	100.0	536	15	AY677221 AY677221 Campyloca
	99	42	100.0	537	6	A57711 A57711 Sequence 21
	100	42	100.0	549	6	AX813676 AX813676 Sequence
	101	42	100.0	549	6	AX813677 AX813677 Sequence
	102	42	100.0	549	6	AX813679 AX813679 Sequence
	103	42	100.0	553	15	AY677222 AY677222 Campyloca
	104	42	100.0	553	15	AY677223 AY677223 Campyloca
	105	42	100.0	553	15	AY677224 AY677224 Campyloca
	106	42	100.0	557	15	AY677225 AY677225 Campyloca
c	107	42	100.0	561	4	AY514428 AY514428 Bubalus b
c	108	42	100.0	569	10	BV412836 BV412836 S229P6385
	109	42	100.0	573	6	AX813675 AX813675 Sequence
	110	42	100.0	580	6	AX813678 AX813678 Sequence
	111	42	100.0	580	6	AX813680 AX813680 Sequence
c	112	42	100.0	587	6	AR568827 AR568827 Sequence
	113	42	100.0	589	6	AX813681 AX813681 Sequence
	114	42	100.0	591	10	BV276443 BV276443 S232P6167
	115	42	100.0	592	6	AX813682 AX813682 Sequence
	116	42	100.0	600	6	AR546074 AR546074 Sequence
	117	42	100.0	601	6	AR664528 AR664528 Sequence
	118	42	100.0	601	6	AR664529 AR664529 Sequence
	119	42	100.0	607	10	BV377419 BV377419 S245P6200
	120	42	100.0	612	6	AR549202 AR549202 Sequence
	121	42	100.0	616	6	AX386212 AX386212 Sequence
	122	42	100.0	619	10	BV229282 BV229282 S233P6299
c	123	42	100.0	620	10	BV268136 BV268136 S235P6208
	124	42	100.0	621	10	BV260649 BV260649 S235P626R
	125	42	100.0	623	10	BV542661 BV542661 G591P5772
c	126	42	100.0	628	10	BV277306 BV277306 S232P6158
	127	42	100.0	632	10	BV305608 BV305608 S236P6321
c	128	42	100.0	641	10	BV061248 BV061248 S212P6042
	129	42	100.0	660	15	CNS019XM AL112226 Botrytis
c	130	42	100.0	668	15	CNS01C21 AL114977 Botrytis
	131	42	100.0	681	10	BV527818 BV527818 G591P6114
c	132	42	100.0	686	10	BV540308 BV540308 G591P6417
	133	42	100.0	687	10	BV541444 BV541444 G591P5583
	134	42	100.0	700	9	AF187066 AF187066 Mus muscu
	135	42	100.0	703	10	BV533882 BV533882 G591P6252
	136	42	100.0	708	6	AR549947 AR549947 Sequence
	137	42	100.0	720	2	AY617399 AY617399 Sterkiell
	138	42	100.0	726	6	AR546135 AR546135 Sequence
c	139	42	100.0	737	10	BV502973 BV502973 qkb91c12.
	140	42	100.0	742	10	BV553142 BV553142 S217P6664
	141	42	100.0	747	6	AR546771 AR546771 Sequence
	142	42	100.0	747	6	AR548753 AR548753 Sequence
	143	42	100.0	768	6	AR552017 AR552017 Sequence
c	144	42	100.0	786	10	PM2C9G AL684736 Penicilli
c	145	42	100.0	789	6	AR556540 AR556540 Sequence
c	146	42	100.0	795	10	BV509645 BV509645 qwf38d09.
c	147	42	100.0	810	10	BV527033 BV527033 G591P6091
	148	42	100.0	820	9	BC027815 BC027815 Mus muscu
c	149	42	100.0	823	10	BV020456 BV020456 S212P6020
	150	42	100.0	837	10	BV618122 BV618122 S217P6186
	151	42	100.0	852	6	AR550376 AR550376 Sequence
	152	42	100.0	854	9	AF097440 AF097440 Mus muscu
	153	42	100.0	858	6	AR501099 AR501099 Sequence
	154	42	100.0	858	6	AR516381 AR516381 Sequence
	155	42	100.0	858	6	AR549587 AR549587 Sequence
c	156	42	100.0	866	10	BV518914 BV518914 G591P6928
	157	42	100.0	867	15	BT020557 BT020557 Arabidops
	158	42	100.0	882	15	AY143847 AY143847 Arabidops
	159	42	100.0	887	9	BC058503 BC058503 Rattus no
	160	42	100.0	888	6	AR548322 AR548322 Sequence
	161	42	100.0	903	6	AR549211 AR549211 Sequence
	162	42	100.0	912	6	A57713 A57713 Sequence 23
	163	42	100.0	921	6	AR551978 AR551978 Sequence
	164	42	100.0	927	6	CQ610476 CQ610476 Sequence
	165	42	100.0	966	6	AR545979 AR545979 Sequence
	166	42	100.0	976	10	CNS06HVC AL399502 T7 end of
c	167	42	100.0	994	10	CNS06HW9 AL399535 T3 end of
c	168	42	100.0	1022	10	CNS06EBE AL395004 T7 end of
	169	42	100.0	1027	15	BT020324 BT020324 Arabidops
	170	42	100.0	1035	6	AR545582 AR545582 Sequence
	171	42	100.0	1065	6	AR545202 AR545202 Sequence
c	172	42	100.0	1087	15	AF457874 AF457874 Euphorbia
	173	42	100.0	1095	15	AK106177 AK106177 Oryza sat
	174	42	100.0	1105	15	AK104498 AK104498 Oryza sat
	175	42	100.0	1105	15	AK104803 AK104803 Oryza sat
c	176	42	100.0	1132	15	AY065142 AY065142 Arabidops
	177	42	100.0	1154	6	AX164109 AX164109 Sequence
	178	42	100.0	1172	6	AX164111 AX164111 Sequence
	179	42	100.0	1179	6	AR550709 AR550709 Sequence
	180	42	100.0	1206	6	CQ803932 CQ803932 Sequence
	181	42	100.0	1233	15	AK068117 AK068117 Oryza sat
	182	42	100.0	1237	15	AY091398 AY091398 Arabidops
	183	42	100.0	1239	5	DQ120961 DQ120961 Hucho tai
c	184	42	100.0	1258	15	AF079247 AF079247 Cymbopogo
	185	42	100.0	1258	15	AF462841 AF462841 Arabidops
	186	42	100.0	1263	6	AR550357 AR550357 Sequence
c	187	42	100.0	1304	15	ZFVICGN Z54365 Z.furfurace
	188	42	100.0	1317	6	AR554685 AR554685 Sequence
	189	42	100.0	1327	15	BT014236 BT014236 Lycopersi
	190	42	100.0	1350	6	AR549628 AR549628 Sequence
c	191	42	100.0	1377	15	AF139533 AF139533 Liquidamb
c	192	42	100.0	1394	15	BT002883 BT002883 Arabidops
	193	42	100.0	1418	15	AK108026 AK108026 Oryza sat
	194	42	100.0	1431	6	AR549760 AR549760 Sequence
c	195	42	100.0	1455	2	PUFFC4 U13883 Bradysia hy
	196	42	100.0	1476	6	AX489036 AX489036 Sequence
	197	42	100.0	1479	6	AR549490 AR549490 Sequence
	198	42	100.0	1479	15	BT004101 BT004101 Arabidops
	199	42	100.0	1482	6	AR547211 AR547211 Sequence
	200	42	100.0	1486	2	DDU72746 U72746 Dictyosteli
	201	42	100.0	1521	6	AR578923 AR578923 Sequence
	202	42	100.0	1521	6	AX240769 AX240769 Sequence
	203	42	100.0	1521	6	AX488745 AX488745 Sequence
c	204	42	100.0	1551	2	AY700082 AY700082 Plasmodiu
	205	42	100.0	1575	6	AR548552 AR548552 Sequence
	206	42	100.0	1610	15	AF005262 AF005262 Cucumis m
	207	42	100.0	1614	6	AR359834 AR359834 Sequence
	208	42	100.0	1616	6	CQ608256 CQ608256 Sequence
	209	42	100.0	1616	15	AY061640 AY061640 Prunus pe
c	210	42	100.0	1632	6	AR553270 AR553270 Sequence
	211	42	100.0	1665	6	AR551237 AR551237 Sequence
	212	42	100.0	1691	15	BT008623 BT008623 Arabidops
	213	42	100.0	1704	15	AK107456 AK107456 Oryza sat
c	214	42	100.0	1716	13	SWP789G M59931 Swinepox vi
c	215	42	100.0	1731	6	AR449702 AR449702 Sequence
	216	42	100.0	1770	6	AR096201 AR096201 Sequence
	217	42	100.0	1770	6	BD080270 BD080270 Transgeni
	218	42	100.0	1770	6	AR262064 AR262064 Sequence
	219	42	100.0	1798	2	DMRSIG1 X63731 D.melanogas
	220	42	100.0	1798	6	AX823825 AX823825 Sequence
	221	42	100.0	1806	6	AR545053 AR545053 Sequence
c	222	42	100.0	1813	6	AR447922 AR447922 Sequence

	223	42	100.0	1853	2	AY128494 AY128494 Drosophil
	224	42	100.0	1897	6	I34914 I34914 Sequence 1
	225	42	100.0	1897	15	YSCOPI1 M57383 Saccharomyc
	226	42	100.0	1898	6	AX536592 AX536592 Sequence
	227	42	100.0	1900	15	AY059751 AY059751 Arabidops
	228	42	100.0	1929	6	CQ582981 CQ582981 Sequence
	229	42	100.0	1949	9	BC029741 BC029741 Mus muscu
	230	42	100.0	1953	6	AR549509 AR549509 Sequence
	231	42	100.0	1958	6	CQ610475 CQ610475 Sequence
	232	42	100.0	1995	6	AR557473 AR557473 Sequence
	233	42	100.0	2000	6	AX509064 AX509064 Sequence
c	234	42	100.0	2000	6	AX509136 AX509136 Sequence
	235	42	100.0	2011	9	AB039933 AB039933 Mus muscu
c	236	42	100.0	2012	2	DROORFB M95826 D.virilis h
	237	42	100.0	2091	6	AR544899 AR544899 Sequence
	238	42	100.0	2127	2	AB210685 AB210685 Ciona int
	239	42	100.0	2152	9	BC083830 BC083830 Rattus no
c	240	42	100.0	2162	15	AK069730 AK069730 Oryza sat
c	241	42	100.0	2212	15	SCYLL010C Z73115 S.cerevisia
c	242	42	100.0	2232	15	CPTKT7 Z46648 C.plantagin
c	243	42	100.0	2270	6	AX461526 AX461526 Sequence
c	244	42	100.0	2347	9	AF191551 AF191551 Mus muscu
	245	42	100.0	2362	2	AK113909 AK113909 Ciona int
	246	42	100.0	2396	2	AF020407 AF020407 Dictyoste
	247	42	100.0	2396	6	AR672184 AR672184 Sequence
	248	42	100.0	2400	2	AF036170 AF036170 Dictyoste
	249	42	100.0	2475	15	AK070717 AK070717 Oryza sat
	250	42	100.0	2506	2	BT001630 BT001630 Drosophil
	251	42	100.0	2541	2	AF339902S2 AF339903 Dictyoste
	252	42	100.0	2555	6	CQ586446 CQ586446 Sequence
	253	42	100.0	2563	2	DMRNAH15 X98766 D.melanogas
	254	42	100.0	2580	6	AR527958 AR527958 Sequence
	255	42	100.0	2606	2	BT022140 BT022140 Drosophil
	256	42	100.0	2663	6	AR439824 AR439824 Sequence
	257	42	100.0	2688	2	AY144597 AY144597 Dictyoste
	258	42	100.0	2690	6	AX073210 AX073210 Sequence
	259	42	100.0	2727	6	AR546931 AR546931 Sequence
c	260	42	100.0	2757	9	MMU95169 U95169 Mus musculu
	261	42	100.0	2790	2	DDI301670 AJ301670 Dictyoste
	262	42	100.0	2838	6	AX488814 AX488814 Sequence
c	263	42	100.0	2842	15	YLI250175 AJ250175 Yarrowia
c	264	42	100.0	2845	15	SCYLR054C Z73226 S.cerevisia
c	265	42	100.0	2859	9	MMU95168 U95168 Mus musculu
	266	42	100.0	2874	2	AY051856 AY051856 Drosophil
c	267	42	100.0	2912	9	MMU95167 U95167 Mus musculu
c	268	42	100.0	2959	6	AX713676 AX713676 Sequence
c	269	42	100.0	2959	8	AK055362 AK055362 Homo sapi
c	270	42	100.0	3014	9	MMU73625 U73625 Mus musculu
	271	42	100.0	3042	6	AX536944 AX536944 Sequence
	272	42	100.0	3089	15	ACH251521 AJ251521 Acremoniu
	273	42	100.0	3159	6	AX058035 AX058035 Sequence
	274	42	100.0	3223	15	YSCSPT8TF M94955 Saccharomyc
	275	42	100.0	3228	6	CQ614937 CQ614937 Sequence
	276	42	100.0	3237	2	DMU89385 U89385 Drosophila
c	277	42	100.0	3277	6	CQ850350 CQ850350 Sequence
c	278	42	100.0	3277	8	AK127474 AK127474 Homo sapi
	279	42	100.0	3317	2	AF145674 AF145674 Drosophil
	280	42	100.0	3336	2	AF111942 AF111942 Dictyoste
	281	42	100.0	3379	2	AF111943 AF111943 Dictyoste
	282	42	100.0	3392	2	AF300334 AF300334 Dictyoste
	283	42	100.0	3455	6	AR146444 AR146444 Sequence
	284	42	100.0	3455	6	AR579950 AR579950 Sequence
	285	42	100.0	3455	15	YSCMPT5 D26184 Yeast MPT5
c	286	42	100.0	3501	6	CS082046 CS082046 Sequence
c	287	42	100.0	3605	15	SCYLR055C Z73227 S.cerevisia
	288	42	100.0	3769	15	YSCHTR1 D25541 Saccharomyc
	289	42	100.0	3815	15	CAALK8 Y14766 Candida alb
	290	42	100.0	3897	15	AF332954 AF332954 Zea mays
c	291	42	100.0	3929	6	CQ582980 CQ582980 Sequence
c	292	42	100.0	3990	2	PUFFC4PR U13892 Bradysia hy
c	293	42	100.0	4048	6	CQ572621 CQ572621 Sequence
c	294	42	100.0	4054	6	CQ589985 CQ589985 Sequence
	295	42	100.0	4146	6	AX489484 AX489484 Sequence
	296	42	100.0	4146	6	AX536824 AX536824 Sequence
	297	42	100.0	4167	15	AF332955 AF332955 Zea mays
	298	42	100.0	4229	6	CQ608255 CQ608255 Sequence
	299	42	100.0	4272	6	CQ578165 CQ578165 Sequence
c	300	42	100.0	4285	6	AR508754 AR508754 Sequence
c	301	42	100.0	4296	5	CT025243 CT025243 Xenopus t
c	302	42	100.0	4492	6	CS124702 CS124702 Sequence
c	303	42	100.0	4565	6	DD013389 DD013389 Methods a
c	304	42	100.0	4565	6	AR542626 AR542626 Sequence
c	305	42	100.0	4565	6	AX235897 AX235897 Sequence
	306	42	100.0	4565	15	ZMAC1 X05424 Maize trans
c	307	42	100.0	4565	15	ZMTAC X01380 Zea mays tr
c	308	42	100.0	4570	15	YLACO5 AJ001303 Yarrowia
	309	42	100.0	4600	15	AF082296 AF082296 Ustilago
	310	42	100.0	4633	15	AF316534 AF316534 Prunus pe
	311	42	100.0	4779	6	CQ596381 CQ596381 Sequence
c	312	42	100.0	4810	15	MZEAC9WX K01964 Maize trans
	313	42	100.0	4911	2	AF275723 AF275723 Dictyoste
	314	42	100.0	4938	6	AX594148 AX594148 Sequence
	315	42	100.0	5015	2	AF076279 AF076279 Dictyoste
	316	42	100.0	5163	6	AR097042 AR097042 Sequence
	317	42	100.0	5163	6	AR279261 AR279261 Sequence
	318	42	100.0	5230	15	AF253047 AF253047 Candida a
	319	42	100.0	5239	6	CQ576536 CQ576536 Sequence
	320	42	100.0	5284	2	DDDOKAGEN X96869 D.discoideu
	321	42	100.0	5318	6	AR097041 AR097041 Sequence
	322	42	100.0	5318	6	AR279260 AR279260 Sequence
	323	42	100.0	5378	15	SCYGL176C Z72698 S.cerevisia
	324	42	100.0	5499	2	AF068065 AF068065 Cryptospo
	325	42	100.0	5501	2	BT010107 BT010107 Drosophil
	326	42	100.0	5511	6	AR097040 AR097040 Sequence
	327	42	100.0	5511	6	AR279259 AR279259 Sequence
c	328	42	100.0	5513	2	AY162292 AY162292 Aplysia c
c	329	42	100.0	5542	6	AX346923 AX346923 Sequence
c	330	42	100.0	5559	6	CQ807300 CQ807300 Sequence
c	331	42	100.0	5571	6	AX251238 AX251238 Sequence
c	332	42	100.0	5571	6	AX251941 AX251941 Sequence
c	333	42	100.0	5571	6	AX344343 AX344343 Sequence
c	334	42	100.0	5571	6	AX348740 AX348740 Sequence
c	335	42	100.0	5621	6	AX345106 AX345106 Sequence
	336	42	100.0	5687	2	AY122262 AY122262 Drosophil
c	337	42	100.0	5728	15	AB167716 AB167716 Rhizopus
	338	42	100.0	5772	15	AF355378 AF355378 Zea mays
c	339	42	100.0	5826	6	AX345465 AX345465 Sequence
	340	42	100.0	6000	6	AR096206 AR096206 Sequence
	341	42	100.0	6000	6	BD080275 BD080275 Transgeni
	342	42	100.0	6000	6	AR262069 AR262069 Sequence
	343	42	100.0	6000	15	ATX98543 X98543 A.thaliana
c	344	42	100.0	6236	6	AX344794 AX344794 Sequence
c	345	42	100.0	6236	6	AX346532 AX346532 Sequence
	346	42	100.0	6289	2	AY160091 AY160091 Dictyoste
	347	42	100.0	6323	15	SCYGL178W Z72700 S.cerevisia
	348	42	100.0	6525	6	AX594150 AX594150 Sequence
c	349	42	100.0	6837	5	MIGCYP1G L41917 Microgadus
	350	42	100.0	6875	11	AF433042 AF433042 Cloning v
	351	42	100.0	6928	11	AF433043 AF433043 Cloning v
c	352	42	100.0	7313	15	ZMU11059 U11059 Zea mays A1
	353	42	100.0	7334	6	AR097039 AR097039 Sequence
	354	42	100.0	7334	6	AR279258 AR279258 Sequence
	355	42	100.0	7487	6	AX594151 AX594151 Sequence
	356	42	100.0	7635	11	AB055064 AB055064 Synthetic
	357	42	100.0	7801	2	AY160101 AY160101 Dictyoste
	358	42	100.0	7916	14	AC014533 AC014533 Drosophil
	359	42	100.0	7951	15	SCDNAGMH X83690 S.cerevisia
	360	42	100.0	7992	14	AC015338 AC015338 Drosophil
c	361	42	100.0	8011	14	AC013239 AC013239 Drosophil
c	362	42	100.0	8167	2	AF474341 AF474341 Dictyoste
	363	42	100.0	8724	4	CDR131945 AJ131945 Camelus d
	364	42	100.0	8867	15	AY191010 AY191010 Zea mays
	365	42	100.0	8979	2	AF411931 AF411931 Plasmodiu
	366	42	100.0	9003	2	AF411932 AF411932 Plasmodiu
	367	42	100.0	9022	2	AF533700 AF533700 Plasmodiu
	368	42	100.0	9024	2	AF411930 AF411930 Plasmodiu

369	42	100.0	9044	2	AF411933	AF411933 Plasmodiu
370	42	100.0	9049	2	AF411929	AF411929 Plasmodiu
c 371	42	100.0	9454	2	BMCHORIO	X15557 Bombyx mori
372	42	100.0	9679	13	CMO580407	AJ580407 Simian im
373	42	100.0	10205	6	CQ596786	CQ596786 Sequence
c 374	42	100.0	10578	14	AC013023	AC013023 Drosophil
375	42	100.0	11123	14	AC013891	AC013891 Drosophil
c 376	42	100.0	11283	1	AF126256	AF126256 Aeromonas
c 377	42	100.0	11401	8	AB209488	AB209488 Homo sapi
378	42	100.0	11587	6	CS133215	CS133215 Sequence
379	42	100.0	13587	6	CQ586445	CQ586445 Sequence
c 380	42	100.0	14287	2	AC115588	AC115588 Dictyoste
381	42	100.0	14720	6	CS018257	CS018257 Sequence
382	42	100.0	14923	14	AC013163	AC013163 Drosophil
383	42	100.0	15983	5	AY458594	AY458594 Ichthyoph
c 384	42	100.0	16987	9	AF362076S1	AF362076 Mus muscu
c 385	42	100.0	17284	9	AL672209	AL672209 Mouse DNA
386	42	100.0	17986	2	U23451	U23451 Caenorhabdi
c 387	42	100.0	17986	2	U23451	U23451 Caenorhabdi
c 388	42	100.0	18158	5	AY277972	AY277972 Takifugu
389	42	100.0	20247	9	AY520825	AY520825 Mus muscu
390	42	100.0	21206	1	AE002104	AE002104 Ureaplasm
c 391	42	100.0	21246	4	BOVCASAS2X	M94327 Bovine alph
392	42	100.0	21693	14	AC017390	AC017390 Drosophil
393	42	100.0	22683	2	AC006656	AC006656 Caenorhab
c 394	42	100.0	22683	2	AC006656	AC006656 Caenorhab
395	42	100.0	23241	14	AC020515	AC020515 Drosophil
c 396	42	100.0	23732	6	CQ614936	CQ614936 Sequence
397	42	100.0	23775	14	AC018178	AC018178 Drosophil
398	42	100.0	24076	6	AX537040	AX537040 Sequence
c 399	42	100.0	25849	14	AC014420	AC014420 Drosophil
400	42	100.0	26857	14	AC013846	AC013846 Drosophil
c 401	42	100.0	28248	14	AC019865	AC019865 Drosophil
c 402	42	100.0	28360	6	CQ578696	CQ578696 Sequence
c 403	42	100.0	28708	11	AY218787	AY218787 Cloning v
c 404	42	100.0	29276	15	YSCH9433	U11582 Saccharomyc
c 405	42	100.0	31034	6	CQ590000	CQ590000 Sequence
c 406	42	100.0	31309	15	AB013387	AB013387 Arabidops
c 407	42	100.0	31361	2	AF016437	AF016437 Caenorhab
408	42	100.0	31901	15	AY292861	AY292861 Boechera
409	42	100.0	32083	15	AF309805	AF309805 Pneumocys
410	42	100.0	32816	6	AX571959	AX571959 Sequence
c 411	42	100.0	32823	8	AC142293	AC142293 Homo sapi
412	42	100.0	33651	2	AC114261	AC114261 Dictyoste
c 413	42	100.0	33780	2	CET05C12	Z66500 Caenorhabdi
c 414	42	100.0	35913	15	AC149414	AC149414 Phakopsor
c 415	42	100.0	36340	15	AC149320	AC149320 Phakopsor
c 416	42	100.0	36708	14	AC020112	AC020112 Drosophil
c 417	42	100.0	36854	14	AC149319	AC149319 Phakopsor
c 418	42	100.0	36933	8	AC107074	AC107074 Homo sapi
c 419	42	100.0	38094	2	AF040647	AF040647 Caenorhab
420	42	100.0	38383	2	CEF35H8	Z36752 Caenorhabdi
421	42	100.0	38449	15	AC152111	AC152111 Pichia st
422	42	100.0	39003	6	AR408742	AR408742 Sequence
423	42	100.0	39003	6	AX067446	AX067446 Sequence
c 424	42	100.0	39074	14	AC156596	AC156596 Laccaria
c 425	42	100.0	39156	15	AC165067	AC165067 Phakopsor
426	42	100.0	39261	14	AC020271	AC020271 Drosophil
c 427	42	100.0	39575	2	AC116925	AC116925 Dictyoste
428	42	100.0	39582	2	DMC17A9	AL009146 Drosophil
c 429	42	100.0	39703	6	AX646691	AX646691 Sequence
c 430	42	100.0	39703	8	AB065476	AB065476 Homo sapi
c 431	42	100.0	40374	5	AL845551	AL845551 Zebrafish
c 432	42	100.0	40690	15	SCLACHXII	X94607 S.cerevisia
c 433	42	100.0	40933	2	DMC140G11	AL035395 Drosophil
434	42	100.0	41511	2	U41559	U41559 Caenorhabdi
435	42	100.0	42273	15	NCB12N19	AL669987 Neurospor
c 436	42	100.0	42642	8	AC006277	AC006277 Homo sapi
c 437	42	100.0	43676	15	SCCEN12RG	X91488 S.cerevisia
c 438	42	100.0	45565	5	FR151J19	AL021531 Fugu rubr
c 439	42	100.0	46165	14	AC136498	AC136498 Drosophil
440	42	100.0	47919	14	AC015044	AC015044 Drosophil
c 441	42	100.0	49628	14	AC132900	AC132900 Homo sapi
c 442	42	100.0	49688	14	AC090162	AC090162 Homo sapi
443	42	100.0	50966	8	AC096992	AC096992 Homo sapi
c 444	42	100.0	51405	14	AC019512	AC019512 Drosophil
445	42	100.0	51413	9	AL845523	AL845523 Mouse DNA
446	42	100.0	51556	8	AL160071	AL160071 Human DNA
447	42	100.0	51975	8	AC108475	AC108475 Homo sapi
c 448	42	100.0	54309	14	AC017353	AC017353 Drosophil
c 449	42	100.0	56310	14	AC084717	AC084717 Homo sapi
c 450	42	100.0	56366	14	AC019800	AC019800 Drosophil
451	42	100.0	56494	2	AE003846_3	Continuation (4 of
c 452	42	100.0	57331	14	AC145220	AC145220 Medicago
453	42	100.0	57743	8	CR759870	CR759870 Human DNA
454	42	100.0	59605	14	AC103926	AC103926 Mus muscu
c 455	42	100.0	59605	14	AC103926	AC103926 Mus muscu
456	42	100.0	60038	14	AC149161	AC149161 Xenopus t
c 457	42	100.0	60038	14	AC149161	AC149161 Xenopus t
c 458	42	100.0	60713	14	AC073922	AC073922 Homo sapi
c 459	42	100.0	60919	14	AC017150	AC017150 Drosophil
460	42	100.0	61052	14	AC123513	AC123513 Dictyoste
c 461	42	100.0	63087	14	AC023446	AC023446 Homo sapi
462	42	100.0	63139	9	AL833790	AL833790 Mouse DNA
c 463	42	100.0	63284	6	CQ600515	CQ600515 Sequence
464	42	100.0	63554	15	H0410G08	AL512546 Oryza sat
465	42	100.0	64000	8	AP002455	AP002455 Homo sapi
c 466	42	100.0	64120	8	AL445646	AL445646 Human DNA
467	42	100.0	64196	8	CR388210	CR388210 Human DNA
c 468	42	100.0	64707	2	AC115607	AC115607 Dictyoste
c 469	42	100.0	64958	14	AC090905	AC090905 Homo sapi
470	42	100.0	65263	14	AC113185	AC113185 Mus muscu
c 471	42	100.0	65572	5	BX927211	BX927211 Zebrafish
472	42	100.0	66075	14	AP007440	AP007440 Lotus cor
c 473	42	100.0	66498	14	AC009199	AC009199 Drosophil
c 474	42	100.0	66874	8	AC069130	AC069130 Homo sapi
475	42	100.0	67847	14	AC017241	AC017241 Drosophil
c 476	42	100.0	68801	8	AP003399	AP003399 Homo sapi
477	42	100.0	69590	15	AC087771	AC087771 Genomic S
478	42	100.0	70534	9	AL929145	AL929145 Mouse DNA
c 479	42	100.0	70670	15	AB024032	AB024032 Arabidops
c 480	42	100.0	71327	15	AP001300	AP001300 Arabidops
c 481	42	100.0	71813	2	DMBN4L24	AL121805 Drosophil
c 482	42	100.0	71962	6	CQ598493	CQ598493 Sequence
483	42	100.0	72006	8	AL607071	AL607071 Human DNA
c 484	42	100.0	72327	2	AC004713	AC004713 Drosophil
485	42	100.0	72409	14	AC131257	AC131257 Homo sapi
486	42	100.0	72974	14	CT005260_7	Continuation (8 of
c 487	42	100.0	73187	14	OSIG00033	AL732338 Oryza sat
c 488	42	100.0	73565	8	AC067962	AC067962 Homo sapi
489	42	100.0	74037	8	HS1071N3	AL031728 Human DNA
490	42	100.0	74509	14	AF260012	AF260012 Homo sapi
491	42	100.0	75148	8	AC091545	AC091545 Homo sapi
c 492	42	100.0	75429	14	AC101416	AC101416 Mus muscu
493	42	100.0	75741	15	AC149204	AC149204 Medicago
c 494	42	100.0	75948	15	AC037424	AC037424 Arabidops
495	42	100.0	76240	14	AC101697	AC101697 Mus muscu
c 496	42	100.0	76336	14	AC023249	AC023249 Homo sapi
497	42	100.0	78404	14	AC147603	AC147603 Zea mays
498	42	100.0	78423	15	AB024035	AB024035 Arabidops
499	42	100.0	78626	15	AC142505	AC142505 Solanum d
500	42	100.0	78908	8	CR788300	CR788300 Human DNA
c 501	42	100.0	79398	14	AC014989	AC014989 Drosophil
c 502	42	100.0	79776	14	AC139431	AC139431 Homo sapi
c 503	42	100.0	80413	15	AC021665	AC021665 Arabidops
504	42	100.0	80518	14	PFMAL13PA	AL109815 Plasmodiu
505	42	100.0	80655	15	AP004577	AP004577 Lotus cor
506	42	100.0	81250	14	AC165439	AC165439 Medicago
507	42	100.0	81358	5	CR382329	CR382329 Zebrafish
c 508	42	100.0	81705	15	AP006384	AP006384 Lotus cor
c 509	42	100.0	82418	14	AP008165	AP008165 Lotus cor
510	42	100.0	83029	15	ATT8O5	AL021890 Arabidops
c 511	42	100.0	83371	15	ATM4I22	AL030978 Arabidops
c 512	42	100.0	84497	9	AC091421	AC091421 Mus muscu
513	42	100.0	84539	14	AC106241_5	Continuation (6 of
c 514	42	100.0	85693	14	AC124117	AC124117 Mus muscu

515	42	100.0	85699	14	AC020021	AC020021 Drosophil
516	42	100.0	86536	15	AP002047	AP002047 Arabidops
c 517	42	100.0	86579	15	AC006587	AC006587 Arabidops
c 518	42	100.0	86869	14	AC151664	AC151664 Medicago
c 519	42	100.0	87230	14	AP007492	AP007492 Lotus cor
520	42	100.0	87767	14	AC014497	AC014497 Drosophil
c 521	42	100.0	87776	14	AP007309	AP007309 Lotus cor
522	42	100.0	87834	8	AP002957	AP002957 Homo sapi
c 523	42	100.0	88549	2	AC116924	AC116924 Dictyoste
524	42	100.0	88709	14	AP007347	AP007347 Lotus cor
525	42	100.0	88929	14	AP007564	AP007564 Lotus cor
526	42	100.0	89275	14	AP007683	AP007683 Lotus cor
c 527	42	100.0	89387	14	AC015216	AC015216 Drosophil
c 528	42	100.0	89426	14	AC166640	AC166640 Solanum d
c 529	42	100.0	89853	14	AP007566	AP007566 Lotus cor
530	42	100.0	90264	14	AP007932	AP007932 Lotus cor
c 531	42	100.0	90275	14	AC127632_3	Continuation (4 of
532	42	100.0	90373	2	AC115680	AC115680 Dictyoste
c 533	42	100.0	90572	14	AC165547	AC165547 Bos tauru
c 534	42	100.0	90621	8	AC092679	AC092679 Homo sapi
c 535	42	100.0	90756	8	AY102285	AY102285 Homo sapi
c 536	42	100.0	91096	14	AC153083	AC153083 Ornithorh
537	42	100.0	91201	8	HSDA86H19	AL049562 Human DNA
c 538	42	100.0	91272	14	AP008143	AP008143 Lotus cor
c 539	42	100.0	92189	15	AC130798	AC130798 Medicago
540	42	100.0	92969	9	BX842667	BX842667 Mouse DNA
c 541	42	100.0	92994	14	AC140159	AC140159 Latimeria
c 542	42	100.0	93305	14	AC138752	AC138752 Homo sapi
543	42	100.0	94175	14	AC136849	AC136849 Rattus no
544	42	100.0	94857	14	AP007998	AP007998 Lotus cor
c 545	42	100.0	94875	14	AC020342	AC020342 Drosophil
546	42	100.0	95546	8	AC022816	AC022816 Homo sapi
c 547	42	100.0	95769	6	AX652128	AX652128 Sequence
c 548	42	100.0	95769	15	AC013430	AC013430 Genomic s
549	42	100.0	96221	15	AC149801	AC149801 Medicago
550	42	100.0	96315	14	AP007568	AP007568 Lotus cor
551	42	100.0	96482	14	AC136041	AC136041 Rattus no
552	42	100.0	97308	14	AC013925	AC013925 Drosophil
553	42	100.0	97460	14	AP008047	AP008047 Lotus cor
c 554	42	100.0	97825	14	AC024642	AC024642 Homo sapi
c 555	42	100.0	98039	8	AY438979	AY438979 Homo sapi
556	42	100.0	98469	15	AP008210_354	Continuation (355
c 557	42	100.0	98760	8	AC133749	AC133749 Homo sapi
558	42	100.0	98942	14	TANN3_18	Continuation (19 o
559	42	100.0	99104	14	AP007508	AP007508 Lotus cor
c 560	42	100.0	99545	8	AC026115	AC026115 Homo sapi
561	42	100.0	99914	5	BX890595	BX890595 Zebrafish
c 562	42	100.0	99967	15	NCB23B10	BX284752 Neurospor
563	42	100.0	100000	8	AP000093	AP000093 Homo sapi
564	42	100.0	100000	8	AP000197	AP000197 Homo sapi
c 565	42	100.0	101021	14	AC161017	AC161017 Pan trogl
c 566	42	100.0	101256	15	CNS07TIW	AL627322 DNA centr
567	42	100.0	101620	8	AC102942	AC102942 Homo sapi
568	42	100.0	101752	15	AP006378	AP006378 Lotus cor
c 569	42	100.0	102004	15	AC005395	AC005395 Arabidops
570	42	100.0	102684	8	BX000688	BX000688 Human DNA
c 571	42	100.0	103011	14	AP008086	AP008086 Lotus cor
572	42	100.0	103979	5	CR352288	CR352288 Zebrafish
573	42	100.0	104423	14	AP008156	AP008156 Lotus cor
574	42	100.0	104775	5	BX469894	BX469894 Zebrafish
c 575	42	100.0	105004	15	AC137065	AC137065 Medicago
c 576	42	100.0	105239	8	AL390959	AL390959 Human DNA
c 577	42	100.0	105297	15	AC126792	AC126792 Medicago
578	42	100.0	105320	2	AC116920	AC116920 Dictyoste
c 579	42	100.0	105472	15	AC150706	AC150706 Medicago
c 580	42	100.0	105682	15	AC148819	AC148819 Medicago
c 581	42	100.0	105768	15	AC069474	AC069474 Arabidops
c 582	42	100.0	106168	8	AC011450	AC011450 Homo sapi
583	42	100.0	106769	14	AP007725	AP007725 Lotus cor
c 584	42	100.0	106806	5	AL935319	AL935319 Zebrafish
c 585	42	100.0	106945	8	AC024337	AC024337 Homo sapi
c 586	42	100.0	107200	15	AC006551	AC006551 Arabidops
c 587	42	100.0	107217	8	AC003070	AC003070 Homo sapi
588	42	100.0	107284	15	AP005290	AP005290 Oryza sat
c 589	42	100.0	107607	8	AP000641	AP000641 Homo sapi
590	42	100.0	107855	14	AP007600	AP007600 Lotus cor
c 591	42	100.0	108100	8	AC103881	AC103881 Homo sapi
592	42	100.0	108582	15	AC011663	AC011663 Arabidops
593	42	100.0	108887	14	AP007707	AP007707 Lotus cor
c 594	42	100.0	109715	14	AC130804	AC130804 Medicago
c 595	42	100.0	109787	15	AC126793	AC126793 Medicago
596	42	100.0	110000	1	AE017225_22	Continuation (23 o
597	42	100.0	110000	1	AE017334_22	Continuation (23 o
598	42	100.0	110000	1	AP006628_6	Continuation (7 of
c 599	42	100.0	110000	1	CP000020_21	Continuation (22 o
c 600	42	100.0	110000	1	CP000082_02	Continuation (3 of
c 601	42	100.0	110000	2	AC116305_0	AC116305 Dictyoste
602	42	100.0	110000	2	AC116305_1	Continuation (2 of
c 603	42	100.0	110000	2	AC116305_1	Continuation (2 of
604	42	100.0	110000	2	AC116305_3	Continuation (4 of
605	42	100.0	110000	2	AC116957_0	AC116957 Dictyoste
c 606	42	100.0	110000	2	AC116957_0	AC116957 Dictyoste
607	42	100.0	110000	2	AC116957_2	Continuation (3 of
c 608	42	100.0	110000	2	AC116984_0	AC116984 Dictyoste
c 609	42	100.0	110000	2	AC116984_1	Continuation (2 of
c 610	42	100.0	110000	2	AC116984_4	Continuation (5 of
c 611	42	100.0	110000	2	AE003426_2	Continuation (3 of
c 612	42	100.0	110000	13	AY318871_1	Continuation (2 of
c 613	42	100.0	110000	14	AC013623_2	Continuation (3 of
614	42	100.0	110000	14	AC106241_4	Continuation (5 of
c 615	42	100.0	110000	14	AC106945_0	AC106945 Rattus no
c 616	42	100.0	110000	14	AC107201_4	Continuation (5 of
c 617	42	100.0	110000	14	AC109729_3	Continuation (4 of
c 618	42	100.0	110000	14	AC110363_4	Continuation (5 of
619	42	100.0	110000	14	AC111637_2	Continuation (3 of
620	42	100.0	110000	14	AC114698_1	Continuation (2 of
c 621	42	100.0	110000	14	AC115160_0	AC115160 Rattus no
622	42	100.0	110000	14	AC118389_4	Continuation (5 of
c 623	42	100.0	110000	14	AC118787_0	AC118787 Rattus no
c 624	42	100.0	110000	14	AC120234_1	Continuation (2 of
c 625	42	100.0	110000	14	AC127632_2	Continuation (3 of
c 626	42	100.0	110000	14	AC140811_0	AC140811 Homo sapi
c 627	42	100.0	110000	14	AC141230_3	Continuation (4 of
c 628	42	100.0	110000	14	AC021632_3	Continuation (4 of
c 629	42	100.0	110000	14	AC153224_0	AC153224 Bos tauru
c 630	42	100.0	110000	14	AC162929_0	AC162929 Mus muscu
c 631	42	100.0	110000	14	AC162929_2	Continuation (3 of
c 632	42	100.0	110000	14	AL359456_1	Continuation (2 of
c 633	42	100.0	110000	14	AL359456_2	Continuation (3 of
c 634	42	100.0	110000	14	AP006494_5	Continuation (6 of
635	42	100.0	110000	14	CR974438_0	CR974438 Mus muscu
c 636	42	100.0	110000	14	CT005250_4	Continuation (5 of
c 637	42	100.0	110000	14	CT005254_4	Continuation (5 of
638	42	100.0	110000	14	CT005259_2	Continuation (3 of
639	42	100.0	110000	14	CT005264_01	Continuation (2 of
c 640	42	100.0	110000	14	CT005265_09	Continuation (10 o
641	42	100.0	110000	14	CT005270_05	Continuation (6 of
642	42	100.0	110000	14	CT009752_4	Continuation (5 of
643	42	100.0	110000	14	CT009752_5	Continuation (6 of
c 644	42	100.0	110000	14	LMFLCHR32_01	Continuation (2 of
645	42	100.0	110000	14	PFMAL13_12	Continuation (13 o
c 646	42	100.0	110000	14	PFMAL8P1_10	Continuation (11 o
c 647	42	100.0	110000	15	AP008214_000	AP008214 Oryza sat
648	42	100.0	110000	15	AP008214_216	Continuation (217
c 649	42	100.0	110000	15	AP008215_111	Continuation (112
c 650	42	100.0	110000	15	AP008216_040	Continuation (41 o
c 651	42	100.0	110000	15	AP008216_053	Continuation (54 o
c 652	42	100.0	110000	15	AP008216_134	Continuation (135
c 653	42	100.0	110000	15	AP008216_135	Continuation (136
c 654	42	100.0	110000	15	AP008217_137	Continuation (138
655	42	100.0	110000	15	AP008217_215	Continuation (216
c 656	42	100.0	110000	15	AP008217_217	Continuation (218
c 657	42	100.0	110000	15	AP008217_228	Continuation (229
658	42	100.0	110000	15	AP008218_018	Continuation (19 o
659	42	100.0	110000	15	AY506529_3	Continuation (4 of
660	42	100.0	110000	15	CR380947_1	Continuation (2 of

	661	42	100.0	110000	15	CR380949_1
c	662	42	100.0	110000	15	CR380953_01
	663	42	100.0	110000	15	CR380953_02
c	664	42	100.0	110000	15	CR380953_03
	665	42	100.0	110000	15	CR380955_06
c	666	42	100.0	110000	15	CR380957_09
c	667	42	100.0	110000	15	CR382123_11
	668	42	100.0	110000	15	CR382124_11
	669	42	100.0	110000	15	CR382126_04
	670	42	100.0	110000	15	CR382126_08
	671	42	100.0	110000	15	CR382126_09
	672	42	100.0	110000	15	CR382126_11
	673	42	100.0	110000	15	CR382127_05
c	674	42	100.0	110000	15	CR382127_10
c	675	42	100.0	110000	15	CR382127_11
	676	42	100.0	110000	15	CR382128_04
	677	42	100.0	110000	15	CR382128_18
c	678	42	100.0	110000	15	CR382129_00
c	679	42	100.0	110000	15	CR382129_25
	680	42	100.0	110000	15	CR382129_31
	681	42	100.0	110000	15	CR382130_35
c	682	42	100.0	110000	15	CR382131_38
c	683	42	100.0	110000	15	CR382132_00
c	684	42	100.0	110000	15	CR382132_16
c	685	42	100.0	110000	15	CR382132_31
c	686	42	100.0	110000	15	CR382135_04
c	687	42	100.0	110000	15	CR382136_02
c	688	42	100.0	110000	15	CR382137_18
	689	42	100.0	110000	15	CR382137_19
	690	42	100.0	110000	15	CR382139_02
c	691	42	100.0	110000	15	AE016820_05
	692	42	100.0	110000	15	AE017342_14
c	693	42	100.0	110000	15	AE017344_16
	694	42	100.0	110000	15	AE017356_1
c	695	42	100.0	110000	15	AP008207_008
	696	42	100.0	110000	15	AP008207_060
c	697	42	100.0	110000	15	AP008207_068
c	698	42	100.0	110000	15	AP008207_083
	699	42	100.0	110000	15	AP008207_101
	700	42	100.0	110000	15	AP008207_221
	701	42	100.0	110000	15	AP008207_222
	702	42	100.0	110000	15	AP008207_346
	703	42	100.0	110000	15	AP008208_306
c	704	42	100.0	110000	15	AP008208_356
	705	42	100.0	110000	15	AP008208_358
c	706	42	100.0	110000	15	AP008209_138
c	707	42	100.0	110000	15	AP008209_139
	708	42	100.0	110000	15	AP008209_221
	709	42	100.0	110000	15	AP008209_255
	710	42	100.0	110000	15	AP008210_064
	711	42	100.0	110000	15	AP008210_310
c	712	42	100.0	110000	15	AP008211_095
c	713	42	100.0	110000	15	AP008211_269
c	714	42	100.0	110000	15	AP008211_270
	715	42	100.0	110000	15	AP008212_061
	716	42	100.0	110000	15	AP008212_191
c	717	42	100.0	110552	14	AC159090
	718	42	100.0	110560	14	AP007372
	719	42	100.0	110578	15	AC156827
	720	42	100.0	110915	15	CR955008
c	721	42	100.0	111067	5	BX663527
c	722	42	100.0	111367	14	AC146818
c	723	42	100.0	111704	14	AC146847
c	724	42	100.0	112057	14	AP004065
	725	42	100.0	112516	8	AC080079
c	726	42	100.0	112525	9	AC153848
c	727	42	100.0	113622	15	AC142526
c	728	42	100.0	113677	8	AL513492
	729	42	100.0	114252	5	BX649441
	730	42	100.0	114377	15	NCB7H23
c	731	42	100.0	114918	15	AP004945
	732	42	100.0	115198	14	AL139223
	733	42	100.0	115223	15	CR940307

Continuation (2 of
Continuation (2 of
Continuation (3 of
Continuation (4 of
Continuation (7 of
Continuation (10 o
Continuation (12 o
Continuation (12 o
Continuation (5 of
Continuation (9 of
Continuation (10 o
Continuation (12 o
Continuation (6 of
Continuation (11 o
Continuation (12 o
Continuation (5 of
Continuation (19 o
CR382129 Yarrowia
Continuation (26 o
Continuation (32 o
Continuation (36 o
Continuation (39 o
CR382132 Yarrowia
Continuation (17 o
Continuation (32 o
Continuation (5 of
Continuation (3 of
Continuation (19 o
Continuation (20 o
Continuation (3 of
Continuation (6 of
Continuation (15 o
Continuation (17 o
Continuation (2 of
Continuation (9 of
Continuation (61 o
Continuation (69 o
Continuation (84 o
Continuation (102
Continuation (222
Continuation (223
Continuation (347
Continuation (307
Continuation (357
Continuation (359
Continuation (139
Continuation (140
Continuation (222
Continuation (256
Continuation (65 o
Continuation (311
Continuation (96 o
Continuation (270
Continuation (271
Continuation (62 o
Continuation (192
AC159090 Medicago
AP007372 Lotus cor
AC156827 Medicago
CR955008 Medicago
BX663527 Chicken D
AC146818 Medicago
AC146847 Ornithorh
AP004065 Oryza sat
AC080079 Homo sapi
AC153848 Mus muscu
AC142526 Medicago
AL513492 Human DNA
BX649441 Zebrafish
BX294026 Neurospor
AP004945 Lotus cor
AL139223 Homo sapi
CR940307 M.truncat

c	734	42	100.0	115359	15	AC003028
c	735	42	100.0	115386	14	AP007401
	736	42	100.0	115804	15	AC124955
c	737	42	100.0	116119	15	AC126781
	738	42	100.0	116803	14	CT010464
	739	42	100.0	116834	15	CR954195
c	740	42	100.0	116975	14	AC010572
c	741	42	100.0	117046	15	AC135600
	742	42	100.0	117338	8	HS173D1
c	743	42	100.0	117526	8	AL390859
c	744	42	100.0	117666	15	AC142507
c	745	42	100.0	118034	4	CR956374
c	746	42	100.0	118106	14	AC080029
	747	42	100.0	118602	14	AP008122
c	748	42	100.0	119044	14	AC018579
c	749	42	100.0	119096	15	AC145021
	750	42	100.0	120194	8	AL355474
c	751	42	100.0	120212	14	AP007468
	752	42	100.0	120458	14	CT009750
c	753	42	100.0	122102	8	AC010468
c	754	42	100.0	122514	5	CR388079
	755	42	100.0	122576	15	NCB13H18
c	756	42	100.0	122624	14	AC015158
c	757	42	100.0	122746	15	CR932958
c	758	42	100.0	122984	14	AC151482
c	759	42	100.0	123160	8	AL353719
c	760	42	100.0	123185	14	AP000587
	761	42	100.0	123272	15	OSJN00099
c	762	42	100.0	123325	15	AP006116
	763	42	100.0	123907	8	AC083851
c	764	42	100.0	125310	15	AC157473
	765	42	100.0	125623	2	AC115599
c	766	42	100.0	125638	8	AC004706
c	767	42	100.0	125679	15	AC027657
	768	42	100.0	125920	15	AP003432
c	769	42	100.0	125958	2	AC115592
c	770	42	100.0	126133	15	AC122146
c	771	42	100.0	127068	15	AC091238
	772	42	100.0	127280	8	CR759766
c	773	42	100.0	127743	14	AP007503
	774	42	100.0	127836	8	AC092590
c	775	42	100.0	127902	15	AP005406
c	776	42	100.0	128361	15	AC137986
	777	42	100.0	128619	15	CNS08CCP
c	778	42	100.0	128831	15	AC148758
	779	42	100.0	128866	14	AP004050
	780	42	100.0	129053	15	AC142096
c	781	42	100.0	129158	4	AC154868
	782	42	100.0	129414	14	AC068398
	783	42	100.0	129517	14	AP007879
c	784	42	100.0	129708	4	AC145407
c	785	42	100.0	129806	8	AC006137
	786	42	100.0	130194	15	AC138087
c	787	42	100.0	130728	15	AP003338
c	788	42	100.0	130735	14	AC134347
	789	42	100.0	130926	2	AC154135
	790	42	100.0	130965	8	AC026773
c	791	42	100.0	130971	15	AC146791
c	792	42	100.0	130971	15	AC147404
	793	42	100.0	131594	14	DMBR17J10
c	794	42	100.0	131860	6	AR659689
	795	42	100.0	131877	8	AC093427
c	796	42	100.0	132719	14	AC166626
c	797	42	100.0	132987	15	AC090874
c	798	42	100.0	133410	5	BX470202
	799	42	100.0	133580	14	AC159613
c	800	42	100.0	133589	15	AP004576
	801	42	100.0	133868	14	AC121106
c	802	42	100.0	134201	14	AC114976
	803	42	100.0	134292	8	HS271M21
c	804	42	100.0	134798	5	AC146502
c	805	42	100.0	135637	8	AP003400
	806	42	100.0	136080	5	AC153487

AC003028 Arabidops
AP007401 Lotus cor
AC124955 Medicago
AC126781 Medicago
CT010464 Medicago
CR954195 Medicago
AC010572 Drosophil
AC135600 Oryza sat
AL031984 Human DNA
AL390859 Human DNA
AC142507 Medicago
CR956374 Pig DNA s
AC080029 Homo sapi
AP008122 Lotus cor
AC018579 Homo sapi
AC145021 Medicago
AL355474 Human DNA
AP007468 Lotus cor
CT009750 Medicago
AC010468 Homo sapi
CR388079 Zebrafish
AL807374 Neurospor
AC015158 Drosophil
CR932958 Medicago
AC151482 Medicago
AL353719 Human DNA
AP000587 Homo sapi
AL606656 Oryza sat
AP006116 Lotus cor
AC083851 Homo sapi
AC157473 Medicago
AC115599 Dictyoste
AC004706 Homo sapi
AC027657 Oryza sat
AP003432 Oryza sat
AC115592 Dictyoste
AC122146 Oryza sat
AC091238 Oryza sat
CR759766 Human DNA
AP007503 Lotus cor
AC092590 Homo sapi
AP005406 Oryza sat
AC137986 Medicago
AL935154 Oryza sat
AC148758 Medicago
AP004050 Oryza sat
AC142096 Medicago
AC154868 Monodelph
AC068398 Homo sapi
AP007879 Lotus cor
AC145407 Macropus
AC006137 Homo sapi
AC138087 Medicago
AP003338 Oryza sat
AC134347 Oryza sat
AC154135 Tribolium
AC026773 Homo sapi
AC146791 Medicago
AC147404 Medicago
AL122026 Drosophil
AR659689 Sequence
AC093427 Homo sapi
AC166626 Sorex ara
AC090874 Oryza sat
BX470202 Zebrafish
AC159613 Medicago
AP004576 Lotus cor
AC121106 Mus muscu
AC114976 Homo sapi
AL031983 Human DNA
AC146502 Danio rer
AP003400 Homo sapi
AC153487 Taeniopyg

	807	42	100.0	136188	14	AC160111 AC160111 Mus muscu
c	808	42	100.0	136240	2	AC117070 AC117070 Dictyoste
	809	42	100.0	136753	14	AC149968 AC149968 Strongylo
	810	42	100.0	136799	15	AP004564 AP004564 Oryza sat
c	811	42	100.0	137718	8	AL137218 AL137218 Human DNA
c	812	42	100.0	138191	15	BX842634 BX842634 Neurospor
	813	42	100.0	139363	9	AC154348 AC154348 Mus muscu
	814	42	100.0	139995	8	AC129416 AC129416 Homo sapi
c	815	42	100.0	140041	15	CR382279 CR382279 M.truncat
c	816	42	100.0	140462	15	AC145219 AC145219 Medicago
c	817	42	100.0	140920	15	AC146947 AC146947 Oryza sat
	818	42	100.0	141189	9	AC156162 AC156162 Mus muscu
c	819	42	100.0	141195	14	AC162388 AC162388 Mus muscu
	820	42	100.0	141331	9	BX005471 BX005471 Mouse DNA
	821	42	100.0	141615	14	AC089989 AC089989 Rattus no
c	822	42	100.0	141715	15	AC130725 AC130725 Oryza sat
	823	42	100.0	142268	15	AP003018 AP003018 Oryza sat
	824	42	100.0	142462	8	AC079238 AC079238 Homo sapi
c	825	42	100.0	142694	8	AC010404 AC010404 Homo sapi
c	826	42	100.0	143353	8	AP000673 AP000673 Homo sapi
c	827	42	100.0	143517	9	AL732418 AL732418 Mouse DNA
c	828	42	100.0	143633	15	AC155340 AC155340 Brassica
c	829	42	100.0	143758	9	AC113265 AC113265 Mus muscu
	830	42	100.0	143842	14	AC134934 AC134934 Bos tauru
c	831	42	100.0	143904	14	AC151656 AC151656 Dasypus n
c	832	42	100.0	143905	14	AC149969 AC149969 Strongylo
	833	42	100.0	143944	8	AC005731 AC005731 Human X B
c	834	42	100.0	144057	14	AC074266 AC074266 Homo sapi
c	835	42	100.0	144092	5	BX465845 BX465845 Zebrafish
	836	42	100.0	144117	8	AC092574 AC092574 Homo sapi
	837	42	100.0	144152	2	AC007725 AC007725 Drosophil
c	838	42	100.0	144322	15	AP002867 AP002867 Oryza sat
	839	42	100.0	144365	2	AC079815 AC079815 Trypanoso
	840	42	100.0	144915	14	CR854923 CR854923 Danio rer
	841	42	100.0	145111	14	AL390245 AL390245 Homo sapi
c	842	42	100.0	145179	8	AL392108 AL392108 Human DNA
c	843	42	100.0	145358	14	AC091086 AC091086 Oryza sat
	844	42	100.0	145431	8	AL662826 AL662826 Human DNA
c	845	42	100.0	146296	14	AC129973 AC129973 Felis cat
	846	42	100.0	146386	15	AC138130 AC138130 Medicago
c	847	42	100.0	146436	9	AL831763 AL831763 Mouse DNA
c	848	42	100.0	146454	13	AF410153 AF410153 Swinepox
	849	42	100.0	146478	14	AC032036 AC032036 Homo sapi
c	850	42	100.0	146570	2	AC117076 AC117076 Dictyoste
c	851	42	100.0	147062	15	AC147405 AC147405 Medicago
	852	42	100.0	147118	8	AC007513 AC007513 Homo sapi
c	853	42	100.0	147138	14	AC157832 AC157832 Strongylo
	854	42	100.0	147207	14	AC078890 AC078890 Oryza sat
	855	42	100.0	147548	15	OSJN00167 AL662965 Oryza sat
	856	42	100.0	147585	15	AP004745 AP004745 Oryza sat
c	857	42	100.0	147588	8	AC098612 AC098612 Homo sapi
c	858	42	100.0	148120	14	AC155517 AC155517 Zea mays
c	859	42	100.0	148600	8	AC002367 AC002367 Homo sapi
	860	42	100.0	148679	8	AC099788 AC099788 Homo sapi
c	861	42	100.0	148764	8	AC148027 AC148027 Homo sapi
	862	42	100.0	148802	2	AC159703 AC159703 Trypanoso
c	863	42	100.0	149139	8	AC010084 AC010084 Homo sapi
c	864	42	100.0	149164	9	AL844558 AL844558 Mouse DNA
c	865	42	100.0	149233	14	AC119665 AC119665 Rattus no
c	866	42	100.0	149253	8	AP002001 AP002001 Homo sapi
c	867	42	100.0	149302	14	AC151876 AC151876 Didelphis
	868	42	100.0	149428	14	AC010264 AC010264 Homo sapi
c	869	42	100.0	149575	14	AC067726 AC067726 Homo sapi
	870	42	100.0	149578	14	AC139820 AC139820 Homo sapi
	871	42	100.0	150032	2	AC159458 AC159458 Trypanoso
c	872	42	100.0	150099	15	AC135916 AC135916 Genomic s
	873	42	100.0	150162	8	BS000046 BS000046 Pan trogl
c	874	42	100.0	150202	5	AL929343 AL929343 Zebrafish
	875	42	100.0	150214	8	AC011291 AC011291 Homo sapi
	876	42	100.0	150551	15	OSJN00088 AL606684 Oryza sat
c	877	42	100.0	150587	2	AC007549 AC007549 Drosophil
	878	42	100.0	150671	2	AC007862 AC007862 Trypanoso
	879	42	100.0	150742	8	AC147649 AC147649 Homo sapi
c	880	42	100.0	151318	14	AL354878 AL354878 Homo sapi
	881	42	100.0	151804	14	AC008161 AC008161 Mus muscu
	882	42	100.0	151806	15	AP002093 AP002093 Oryza sat
c	883	42	100.0	152327	14	AC026295 AC026295 Homo sapi
	884	42	100.0	152506	9	AC145589 AC145589 Mus muscu
c	885	42	100.0	152532	9	AC126429 AC126429 Mus muscu
c	886	42	100.0	152594	8	AL158196 AL158196 Human DNA
	887	42	100.0	152595	14	AC140167 AC140167 Homo sapi
c	888	42	100.0	152785	14	CR450771 CR450771 Danio rer
c	889	42	100.0	152828	15	AC133859 AC133859 Oryza sat
c	890	42	100.0	153037	14	BX927246 BX927246 Danio rer
c	891	42	100.0	153239	15	AC120887 AC120887 Oryza sat
c	892	42	100.0	153265	8	AP000237 AP000237 Homo sapi
	893	42	100.0	153751	2	AC116551 AC116551 Dictyoste
c	894	42	100.0	153751	2	AC116551 AC116551 Dictyoste
c	895	42	100.0	154034	9	AL772336 AL772336 Mouse DNA
	896	42	100.0	154071	2	AC115598 AC115598 Dictyoste
	897	42	100.0	154158	8	AC015600 AC015600 Homo sapi
	898	42	100.0	154248	15	AP003631 AP003631 Oryza sat
c	899	42	100.0	154311	8	AF107258 AF107258 Homo sapi
c	900	42	100.0	154681	2	AC099008 AC099008 Drosophil
	901	42	100.0	154799	6	CQ926648 CQ926648 Sequence
	902	42	100.0	155210	8	AC007447 AC007447 Homo sapi
	903	42	100.0	155254	2	AC007173 AC007173 Drosophil
	904	42	100.0	155397	8	AL590523 AL590523 Human DNA
c	905	42	100.0	155429	9	AC153886 AC153886 Mus muscu
c	906	42	100.0	155766	14	AC147016 AC147016 Zea mays
	907	42	100.0	155874	8	AL645936 AL645936 Human DNA
c	908	42	100.0	155983	8	AY884207 AY884207 Homo sapi
	909	42	100.0	156060	14	AC004153 AC004153 Plasmodiu
	910	42	100.0	156392	14	AC159303 AC159303 Mus muscu
	911	42	100.0	156415	8	AC011464 AC011464 Homo sapi
c	912	42	100.0	156899	5	BX465856 BX465856 Zebrafish
c	913	42	100.0	157087	14	AC142155 AC142155 Rattus no
	914	42	100.0	157118	8	AC069286 AC069286 Homo sapi
	915	42	100.0	157141	8	AC011470 AC011470 Homo sapi
	916	42	100.0	157193	8	AL355315 AL355315 Human DNA
c	917	42	100.0	157226	8	AC018902 AC018902 Homo sapi
	918	42	100.0	157233	9	AC101714 AC101714 Mus muscu
	919	42	100.0	157289	14	AC021447 AC021447 Homo sapi
c	920	42	100.0	157301	2	AC159409 AC159409 Trypanoso
c	921	42	100.0	157330	8	AP000872 AP000872 Homo sapi
c	922	42	100.0	157531	8	AL133409 AL133409 Human DNA
c	923	42	100.0	157565	14	AC161533 AC161533 Mus muscu
	924	42	100.0	157572	8	AL513210 AL513210 Human DNA
c	925	42	100.0	157579	8	AC148030 AC148030 Homo sapi
c	926	42	100.0	157583	5	BX571951 BX571951 Zebrafish
	927	42	100.0	157672	8	AC113143 AC113143 Homo sapi
	928	42	100.0	157871	8	AC104365 AC104365 Homo sapi
	929	42	100.0	157881	14	CR936357 CR936357 Danio rer
	930	42	100.0	158482	13	AY126275 AY126275 Mamestra
c	931	42	100.0	158497	14	AC139517 AC139517 Rattus no
	932	42	100.0	158801	14	AC128036 AC128036 Rattus no
	933	42	100.0	158807	14	AC011771 AC011771 Homo sapi
	934	42	100.0	158857	15	AP005470 AP005470 Oryza sat
c	935	42	100.0	158913	9	AL645928 AL645928 Mouse DNA
	936	42	100.0	158952	14	CR550301 CR550301 Danio rer
	937	42	100.0	159064	8	AC097516 AC097516 Homo sapi
c	938	42	100.0	159167	14	DMBR42L18 AL122031 Drosophil
c	939	42	100.0	159190	14	AC150376 AC150376 Canis fam
c	940	42	100.0	159242	15	AP002902 AP002902 Oryza sat
	941	42	100.0	159513	2	AC010706 AC010706 Drosophil
	942	42	100.0	159914	8	AC027553 AC027553 Homo sapi
c	943	42	100.0	160042	14	AL365211 AL365211 Homo sapi
c	944	42	100.0	160374	5	CR759918 CR759918 Zebrafish
c	945	42	100.0	160410	15	AP005685 AP005685 Oryza sat
c	946	42	100.0	160444	2	AC010123 AC010123 Drosophil
c	947	42	100.0	160466	14	AC139524 AC139524 Homo sapi
	948	42	100.0	160484	15	OSJN00255 AL731610 Oryza sat
	949	42	100.0	160496	14	AC164419 AC164419 Mus muscu
	950	42	100.0	160498	2	AC007864 AC007864 Trypanoso
	951	42	100.0	160518	14	AC149942 AC149942 Strongylo
c	952	42	100.0	160601	8	AC021107 AC021107 Homo sapi

	953	42	100.0	160634	14	AC141999 AC141999 Rattus no
c	954	42	100.0	160655	8	AC064874 AC064874 Homo sapi
	955	42	100.0	160746	14	AC044853 AC044853 Homo sapi
c	956	42	100.0	160792	9	BX682533 BX682533 Mouse DNA
	957	42	100.0	161081	5	CR855273 CR855273 Zebrafish
	958	42	100.0	161106	14	AP001133 AP001133 Homo sapi
c	959	42	100.0	161206	14	AC141737 AC141737 Apis mell
c	960	42	100.0	161266	15	AP001278 AP001278 Oryza sat
c	961	42	100.0	161274	14	AC139613 AC139613 Rattus no
	962	42	100.0	161452	8	AC020925 AC020925 Homo sapi
	963	42	100.0	161523	14	AC148427 AC148427 Macropus
	964	42	100.0	161608	14	AC140835 AC140835 Homo sapi
	965	42	100.0	161676	15	AC135907 AC135907 Genomic s
	966	42	100.0	161701	14	AC117110 AC117110 Rattus no
c	967	42	100.0	161800	14	CR847528 CR847528 Danio rer
c	968	42	100.0	161850	8	AF306443 AF306443 Homo sapi
	969	42	100.0	161905	9	AL929034 AL929034 Mouse DNA
	970	42	100.0	161940	14	AC148232 AC148232 Didelphis
c	971	42	100.0	161965	9	AC148008 AC148008 Mus muscu
c	972	42	100.0	161972	8	AP001985 AP001985 Homo sapi
	973	42	100.0	162631	14	BX950865 BX950865 Danio rer
	974	42	100.0	162692	14	AC016171 AC016171 Homo sapi
c	975	42	100.0	162949	2	AC008228 AC008228 Drosophil
c	976	42	100.0	163087	14	AC139383 AC139383 Rattus no
c	977	42	100.0	163300	14	AC023369 AC023369 Homo sapi
	978	42	100.0	163307	2	AC011703 AC011703 Drosophil
c	979	42	100.0	163384	8	AC026409 AC026409 Homo sapi
c	980	42	100.0	163521	9	AC121966 AC121966 Mus muscu
c	981	42	100.0	163690	14	AC131509 AC131509 Strongylo
	982	42	100.0	163772	8	AC093165 AC093165 Homo sapi
	983	42	100.0	164275	14	AC021634 AC021634 Homo sapi
c	984	42	100.0	164304	14	AP002398 AP002398 Homo sapi
	985	42	100.0	164631	14	AC163335 AC163335 Mus muscu
c	986	42	100.0	164663	9	AC153870 AC153870 Mus muscu
c	987	42	100.0	164679	15	AC078944 AC078944 Genomic S
c	988	42	100.0	164834	9	AL929043 AL929043 Mouse DNA
c	989	42	100.0	164873	2	AC159405 AC159405 Trypanoso
	990	42	100.0	164884	2	AC011063 AC011063 Drosophil
	991	42	100.0	165004	14	AC139465 AC139465 Homo sapi
c	992	42	100.0	165278	9	AC153959 AC153959 Mus muscu
	993	42	100.0	165366	14	AP001789 AP001789 Homo sapi
	994	42	100.0	165501	14	AC037468 AC037468 Homo sapi
	995	42	100.0	165550	15	AP003488 AP003488 Oryza sat
	996	42	100.0	165682	14	AC163716 AC163716 Pan trogl
c	997	42	100.0	165682	14	AC163716 AC163716 Pan trogl
c	998	42	100.0	165738	14	AC019165 AC019165 Homo sapi
	999	42	100.0	165892	5	AL929516 AL929516 Zebrafish
	c1000	42	100.0	165977	14	CR933539 CR933539 Danio rer

ALIGNMENTS

```
RESULT 1
AR579290/c
LOCUS       AR579290                21 bp    DNA     linear   PAT 14-DEC-2004
DEFINITION  Sequence 435 from patent US 6783985.
ACCESSION   AR579290
VERSION     AR579290.1  GI:56582107
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 21)
  AUTHORS   Roemer,T., Jiang,B., Boone,C. and Bussey,H.
  TITLE     Gene disruption methodologies for drug target discovery
  JOURNAL   Patent: US 6783985-A 435 31-AUG-2004;
            Elitra Pharmaceuticals Inc.; San Diego, CA
FEATURES             Location/Qualifiers
     source          1. .21
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:                 5.67          Length:          21
Score:                     42.00         Matches:         6
Percent Similarity:        100.00%       Conservative:    0
Best Local Similarity:     100.00%       Mismatches:      0
Query Match:               100.00%       Indels:          0
DB:                        6             Gaps:            0

US-09-858-332G-28 (1-6) x AR579290 (1-21)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         20 CACAACCACAACCACAAC 3

RESULT 2
AX241197/c
LOCUS       AX241197                21 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 435 from Patent WO0160975.
ACCESSION   AX241197
VERSION     AX241197.1  GI:15798072
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Roemer,T., Jiang,B., Boone,C. and Bussey,H.
  TITLE     Gene disruption methodologies for drug target discovery
  JOURNAL   Patent: WO 0160975-A 435 23-AUG-2001;
            Elitra Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .21
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="DNA primer"
ORIGIN

Alignment Scores:
Pred. No.:                 5.67          Length:          21
Score:                     42.00         Matches:         6
Percent Similarity:        100.00%       Conservative:    0
Best Local Similarity:     100.00%       Mismatches:      0
Query Match:               100.00%       Indels:          0
DB:                        6             Gaps:            0

US-09-858-332G-28 (1-6) x AX241197 (1-21)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         20 CACAACCACAACCACAAC 3

RESULT 3
AX487745/c
LOCUS       AX487745                21 bp    DNA     linear   PAT 16-AUG-2002
DEFINITION  Sequence 5045 from Patent WO02053728.
ACCESSION   AX487745
VERSION     AX487745.1  GI:22321825
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1
  AUTHORS   Roemer,T., Jiang,B., Boone,C., Bussey,H. and Ohlsen,K.L.
  TITLE     Gene disruption methodologies for drug target discovery
  JOURNAL   Patent: WO 02053728-A 5045 11-JUL-2002;
            Elitra Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .21
                     /organism="Candida albicans"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:5476"
ORIGIN
```

```
Alignment Scores:
Pred. No.:              5.67          Length:        21
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-28 (1-6) x AX487745 (1-21)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         20 CACAACCACAACCACAAC 3

RESULT 4
AX594160
LOCUS       AX594160                  42 bp    DNA     linear   PAT 13-FEB-2003
DEFINITION  Sequence 13 from Patent WO02083910.
ACCESSION   AX594160
VERSION     AX594160.1  GI:28375390
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Farmer,A.A.
  TITLE     Sequence specific recombinase-based methods for producing intron
            containing vectors and compositions for use in practicing the same
  JOURNAL   Patent: WO 02083910-A 13 24-OCT-2002;
            Clontech Laboratories Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .42
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="splice sequence"
ORIGIN

Alignment Scores:
Pred. No.:              11            Length:        42
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-28 (1-6) x AX594160 (1-42)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db          4 CATAATCATAATCATAAT 21

RESULT 5
CQ754680/c
LOCUS       CQ754680                  57 bp    DNA     linear   PAT 01-MAR-2004
DEFINITION  Sequence 5 from Patent WO2003062415.
ACCESSION   CQ754680
VERSION     CQ754680.1  GI:44845818
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Drabek,D., Dekker,S. and Grosveld,F.G.
  TITLE     Transgenic animal and methods
  JOURNAL   Patent: WO 2003062415-A 5 31-JUL-2003;
            Erasmus Unviersiteit (NL)
FEATURES             Location/Qualifiers
     source          1. .57
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="PCR primer"
ORIGIN

Alignment Scores:
Pred. No.:              14.8          Length:        57
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-28 (1-6) x CQ754680 (1-57)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         38 CACAACCACAACCACAAC 21

RESULT 6
AX813672/c
LOCUS       AX813672                  57 bp    DNA     linear   PAT 02-DEC-2003
DEFINITION  Sequence 5 from Patent WO03062415.
ACCESSION   AX813672
VERSION     AX813672.1  GI:38636040
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Drabek,D., Dekker,S. and Grosveld,F.G.
  TITLE     Transgenic animal and methods
  JOURNAL   Patent: WO 03062415-A 5 31-JUL-2003;
            Erasmus Unviersiteit (NL)
FEATURES             Location/Qualifiers
     source          1. .57
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="PCR primer"
ORIGIN

Alignment Scores:
Pred. No.:              14.8          Length:        57
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-28 (1-6) x AX813672 (1-57)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         38 CACAACCACAACCACAAC 21

RESULT 7
AX483190
LOCUS       AX483190                  65 bp    DNA     linear   PAT 16-AUG-2002
DEFINITION  Sequence 490 from Patent WO02053728.
ACCESSION   AX483190
VERSION     AX483190.1  GI:22317610
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1
  AUTHORS   Roemer,T., Jiang,B., Boone,C., Bussey,H. and Ohlsen,K.L.
  TITLE     Gene disruption methodologies for drug target discovery
  JOURNAL   Patent: WO 02053728-A 490 11-JUL-2002;
            Elitra Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .65
                     /organism="Candida albicans"
```

```
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:5476"
ORIGIN

Alignment Scores:
Pred. No.:                16.8          Length:         65
Score:                    42.00         Matches:        6
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              100.00%       Indels:         0
DB:                       6             Gaps:           0

US-09-858-332G-28 (1-6) x AX483190 (1-65)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         47 CATAACCATAACCATAAC 64

RESULT 8
AX485536
LOCUS       AX485536                 65 bp    DNA     linear   PAT 16-AUG-2002
DEFINITION  Sequence 2836 from Patent WO02053728.
ACCESSION   AX485536
VERSION     AX485536.1  GI:22319820
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1
  AUTHORS   Roemer,T., Jiang,B., Boone,C., Bussey,H. and Ohlsen,K.L.
  TITLE     Gene disruption methodologies for drug target discovery
  JOURNAL   Patent: WO 02053728-A 2836 11-JUL-2002;
            Elitra Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .65
                     /organism="Candida albicans"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:5476"
ORIGIN

Alignment Scores:
Pred. No.:                16.8          Length:         65
Score:                    42.00         Matches:        6
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              100.00%       Indels:         0
DB:                       6             Gaps:           0

US-09-858-332G-28 (1-6) x AX485536 (1-65)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db          5 CATAACCACAACCACAAC 22

RESULT 9
AX594158
LOCUS       AX594158                 100 bp    DNA     linear   PAT 13-FEB-2003
DEFINITION  Sequence 11 from Patent WO02083910.
ACCESSION   AX594158
VERSION     AX594158.1  GI:28375388
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Farmer,A.A.
  TITLE     Sequence specific recombinase-based methods for producing intron
            containing vectors and compositions for use in practicing the same
  JOURNAL   Patent: WO 02083910-A 11 24-OCT-2002;
            Clontech Laboratories Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .100
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="splice sequence"
ORIGIN

Alignment Scores:
Pred. No.:                25.5          Length:         100
Score:                    42.00         Matches:        6
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              100.00%       Indels:         0
DB:                       6             Gaps:           0

US-09-858-332G-28 (1-6) x AX594158 (1-100)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         61 CATAATCATAATCATAAT 78

RESULT 10
AY559840/c
LOCUS       AY559840                 115 bp    DNA     linear   MAM 19-MAR-2004
DEFINITION  Bos taurus germline immunoglobulin heavy chain gene, DH3 segment,
            partial sequence.
ACCESSION   AY559840
VERSION     AY559840.1  GI:45551582
KEYWORDS    .
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 115)
  AUTHORS   Shojaei,F., Saini,S.S. and Kaushik,A.K.
  TITLE     Unusually long germline DH genes contribute to large sized CDR3H in
            bovine antibodies
  JOURNAL   Mol. Immunol. 40 (1), 61-67 (2003)
   PUBMED   12909131
REFERENCE   2  (bases 1 to 115)
  AUTHORS   Shojaei,F., Saini,S.S. and Kaushik,A.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-FEB-2004) Pathobiology, University of Guelph,
            University of Guelph, Guelph, ON N1G 2W1, Canada
FEATURES             Location/Qualifiers
     source          1. .115
                     /organism="Bos taurus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9913"
                     /chromosome="21"
                     /map="21q24"
                     /germline
     gene            <1. .>115
                     /gene="immunoglobulin heavy chain"
     D_segment       1. .115
                     /gene="immunoglobulin heavy chain"
                     /note="DH3"
     repeat_region   1. .9
                     /note="recombination signal sequence"
     repeat_region   23. .29
                     /note="recombination signal sequence"
     repeat_region   88. .94
                     /note="recombination signal sequence"
     repeat_region   107. .115
                     /note="recombination signal sequence"
ORIGIN

Alignment Scores:
Pred. No.:                29.1          Length:         115
Score:                    42.00         Matches:        6
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              100.00%       Indels:         0
```

```
DB:                    4              Gaps:          0

US-09-858-332G-28 (1-6) x AY559840 (1-115)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         80 CATAACCATAACCATAAC 63

RESULT 11
AR568829/c
LOCUS       AR568829                 115 bp    DNA     linear   PAT 14-DEC-2004
DEFINITION  Sequence 3 from patent US 6740747.
ACCESSION   AR568829
VERSION     AR568829.1  GI:56568471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 115)
  AUTHORS   Kaushik,A. and Saini,S.S.
  TITLE     Bovine VDJ cassette, BF1H1, suitable for antigenization
  JOURNAL   Patent: US 6740747-A 3 25-MAY-2004;
FEATURES             Location/Qualifiers
     source          1. .115
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              29.1          Length:        115
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-28 (1-6) x AR568829 (1-115)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         80 CATAACCATAACCATAAC 63

RESULT 12
AJ811567
LOCUS       AJ811567                 122 bp    mRNA    linear   MAM 13-OCT-2004
DEFINITION  Lama glama partial mRNA for immunoglobulin heavy chain hinge region
            (ighc gene), clone pRL188.
ACCESSION   AJ811567
VERSION     AJ811567.1  GI:51490834
KEYWORDS    ighc gene; immunoglobulin heavy chain hinge region.
SOURCE      Lama glama (llama)
  ORGANISM  Lama glama
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Tylopoda;
            Camelidae; Lama.
REFERENCE   1
  AUTHORS   Harmsen,M.M., Van Solt,C.B. and Van Keulen,L.
  TITLE     Selection of llama single-domain antibody fragments binding to
            recombinant bovine prion protein in ELISA
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 122)
  AUTHORS   Harmsen,M.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-AUG-2004) Harmsen M.M., Infection Biology, Animal
            Sciences Group (ID-Lelystad BV), Edelhertweg 15, Lelystad, 8219 PH,
            NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .122
                     /organism="Lama glama"
                     /mol_type="mRNA"
                     /db_xref="taxon:9844"
                     /clone="pRL188"
                     /cell_type="B lymphocyte"
                     /rearranged
     gene            1. .122
                     /gene="ighc"
     primer_bind     <1. .25
                     /gene="ighc"
                     /note="primer LH derived"
     CDS             <17. .>93
                     /gene="ighc"
                     /codon_start=1
                     /product="immunoglobulin heavy chain hinge region"
                     /protein_id="CAH18863.1"
                     /db_xref="GI:51490835"
                     /translation="EPKTPKPQPQPQPQPQPNPTTESKCP"
     primer_bind     71. .>122
                     /gene="ighc"
                     /note="primer BOLI269 derived"
ORIGIN

Alignment Scores:
Pred. No.:              30.9          Length:        122
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     4             Gaps:          0

US-09-858-332G-28 (1-6) x AJ811567 (1-122)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         36 CACAACCACAACCACAAC 53

RESULT 13
AR568831/c
LOCUS       AR568831                 139 bp    DNA     linear   PAT 14-DEC-2004
DEFINITION  Sequence 5 from patent US 6740747.
ACCESSION   AR568831
VERSION     AR568831.1  GI:56568473
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 139)
  AUTHORS   Kaushik,A. and Saini,S.S.
  TITLE     Bovine VDJ cassette, BF1H1, suitable for antigenization
  JOURNAL   Patent: US 6740747-A 5 25-MAY-2004;
FEATURES             Location/Qualifiers
     source          1. .139
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              35            Length:        139
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-28 (1-6) x AR568831 (1-139)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         50 CATAACCATAACCACAAC 33

RESULT 14
CNS01B16/c
LOCUS       CNS01B16                 180 bp    mRNA    linear   PLN 20-MAR-2004
DEFINITION  Botrytis cinerea strain T4 cDNA library.
ACCESSION   AL113650
VERSION     AL113650.1  GI:5828269
KEYWORDS    cDNA library; EST, nitrate deprivation.
```

```
SOURCE      Botryotinia fuckeliana
  ORGANISM  Botryotinia fuckeliana
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomycetes;
            Helotiales; Sclerotiniaceae; Botryotinia.
REFERENCE   1  (bases 1 to 180)
  AUTHORS   Bitton,F., Levis,C., Fortini,D., Pradier,J.M. and Brygoo,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-1999) Phytopathologie, INRA, route de St Cyr,
            78026 Versailles, France
REFERENCE   2  (bases 1 to 180)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-1999) Genoscope - Centre National de Sequencage :
            CP 5706 91057 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     The cDNA library was made by using Botrytis cinerea T4 strain grown
            under nitrate deprivation conditions. The cDNA were cloned
            directionally in the pBSII vector.
FEATURES             Location/Qualifiers
     source          1. .180
                     /organism="Botryotinia fuckeliana"
                     /mol_type="mRNA"
                     /strain="T4"
                     /db_xref="taxon:40559"
                     /note="Genoscope sequence ID : W13F091"
ORIGIN

Alignment Scores:
Pred. No.:                 44.9          Length:          180
Score:                     42.00         Matches:         6
Percent Similarity:        100.00%       Conservative:    0
Best Local Similarity:     100.00%       Mismatches:      0
Query Match:               100.00%       Indels:          0
DB:                        15            Gaps:            0

US-09-858-332G-28 (1-6) x CNS01B16 (1-180)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         25 CATAACCATAACCATAAC 8

RESULT 15
BTA512824/c
LOCUS       BTA512824               190 bp    mRNA    linear   MAM 28-JUL-2004
DEFINITION  Bos taurus partial mRNA for immunoglobulin heavy chain variable
            region (IGHV gene).
ACCESSION   AJ512824
VERSION     AJ512824.1  GI:24412782
KEYWORDS    IGHV gene; immunoglobulin heavy chain; variable region.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1
  AUTHORS   Panelli,S., Damiani,G., Galli,C. and Sgaramella,V.
  TITLE     Rearranged genomes of bovine blood cells can allow the development
            of clones till late fetal stages; but rare unrearranged genomes
            have greater potential and lead to adulthood
  JOURNAL   Gene 334, 99-103 (2004)
  PUBMED    15256259
REFERENCE   2  (bases 1 to 190)
  AUTHORS   Panelli,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2002) Panelli S., Patol. Anim. Igiene e Sanita
            Pubbl. Vet., Fondazione PTP c/o Universita di Milano, Via Celoria
            10, Milano 20133, ITALY
FEATURES             Location/Qualifiers
     source          1. .190
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /cell_type="lymphocytes"
                     /rearranged
     gene            1. .190
                     /gene="IGHV"
     CDS             <1. .>190
                     /gene="IGHV"
                     /codon_start=2
                     /product="immunoglobulin heavy chain variable region"
                     /protein_id="CAD55149.1"
                     /db_xref="GI:24412783"
                     /translation="PGLKSRLSITKDNSKSQVFLSLSSLTPEDTATYYCASMVVMVMV
                     VMVVMVMVMVMVYVDAWGQ"
     V_region        <1. .>190
                     /gene="IGHV"
     V_segment       1. .97
                     /gene="IGHV"
     J_segment       98. .190
                     /gene="IGHV"
ORIGIN

Alignment Scores:
Pred. No.:                 47.3          Length:          190
Score:                     42.00         Matches:         6
Percent Similarity:        100.00%       Conservative:    0
Best Local Similarity:     100.00%       Mismatches:      0
Query Match:               100.00%       Indels:          0
DB:                        4             Gaps:            0

US-09-858-332G-28 (1-6) x BTA512824 (1-190)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        166 CATAACCATAACCATAAC 149

RESULT 16
AR553849
LOCUS       AR553849                192 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 8980 from patent US 6747137.
ACCESSION   AR553849
VERSION     AR553849.1  GI:53947024
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 192)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 8980 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .192
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:                 47.7          Length:          192
Score:                     42.00         Matches:         6
Percent Similarity:        100.00%       Conservative:    0
Best Local Similarity:     100.00%       Mismatches:      0
Query Match:               100.00%       Indels:          0
DB:                        6             Gaps:            0

US-09-858-332G-28 (1-6) x AR553849 (1-192)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        120 CACAACCACAACCACAAC 137

RESULT 17
AY559839/c
LOCUS       AY559839                205 bp    DNA     linear   MAM 19-MAR-2004
DEFINITION  Bos taurus germline immunoglobulin heavy chain gene, DH2 segment,
```

```
            partial sequence.
ACCESSION   AY559839
VERSION     AY559839.1  GI:45551581
KEYWORDS    .
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 205)
  AUTHORS   Shojaei,F., Saini,S.S. and Kaushik,A.K.
  TITLE     Unusually long germline DH genes contribute to large sized CDR3H in
            bovine antibodies
  JOURNAL   Mol. Immunol. 40 (1), 61-67 (2003)
   PUBMED   12909131
REFERENCE   2  (bases 1 to 205)
  AUTHORS   Shojaei,F., Saini,S.S. and Kaushik,A.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-FEB-2004) Pathobiology, University of Guelph,
            University of Guelph, Guelph, ON N1G 2W1, Canada
FEATURES             Location/Qualifiers
     source          1. .205
                     /organism="Bos taurus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9913"
                     /chromosome="21"
                     /map="21q24"
                     /germline
     gene            <1. .>205
                     /gene="immunoglobulin heavy chain"
     D_segment       1. .205
                     /gene="immunoglobulin heavy chain"
                     /note="DH2"
     repeat_region   1. .9
                     /note="recombination signal sequence"
     repeat_region   23. .29
                     /note="recombination signal sequence"
     repeat_region   177. .184
                     /note="recombination signal sequence"
     repeat_region   197. .205
                     /note="recombination signal sequence"
ORIGIN

Alignment Scores:
Pred. No.:              50.9          Length:         205
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     4             Gaps:           0

US-09-858-332G-28 (1-6) x AY559839 (1-205)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         80 CATAACCATAACCACAAC 63

RESULT 18
AR568828/c
LOCUS       AR568828                205 bp    DNA     linear   PAT 14-DEC-2004
DEFINITION  Sequence 2 from patent US 6740747.
ACCESSION   AR568828
VERSION     AR568828.1  GI:56568470
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 205)
  AUTHORS   Kaushik,A. and Saini,S.S.
  TITLE     Bovine VDJ cassette, BF1H1, suitable for antigenization
  JOURNAL   Patent: US 6740747-A 2 25-MAY-2004;
FEATURES             Location/Qualifiers
     source          1. .205
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              50.9          Length:         205
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     6             Gaps:           0

US-09-858-332G-28 (1-6) x AR568828 (1-205)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         80 CATAACCATAACCACAAC 63

RESULT 19
BTA512976/c
LOCUS       BTA512976               209 bp    DNA     linear   MAM 28-JUL-2004
DEFINITION  Bos taurus partial IGHV gene for immunoglobulin heavy chain
            variable region, exon 1, isolate adult 1, clone 2.
ACCESSION   AJ512976
VERSION     AJ512976.1  GI:24459870
KEYWORDS    IGHV gene; immunoglobulin heavy chain; variable region.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1
  AUTHORS   Panelli,S., Damiani,G., Galli,C. and Sgaramella,V.
  TITLE     Rearranged genomes of bovine blood cells can allow the development
            of clones till late fetal stages; but rare unrearranged genomes
            have greater potential and lead to adulthood
  JOURNAL   Gene 334, 99-103 (2004)
   PUBMED   15256259
REFERENCE   2  (bases 1 to 209)
  AUTHORS   Panelli,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT-2002) Panelli S., Patol. Anim. Igiene e Sanita
            Pubbl. Vet., Fondazione PTP c/o Universita di Milano, via Celoria
            10, Milano 20133, ITALY
FEATURES             Location/Qualifiers
     source          1. .209
                     /organism="Bos taurus"
                     /mol_type="genomic DNA"
                     /isolate="adult 1"
                     /db_xref="taxon:9913"
                     /clone="2"
                     /cell_type="lymphocytes"
                     /dev_stage="adult"
                     /rearranged
     gene            1. .209
                     /gene="IGHV"
     CDS             <1. .>209
                     /gene="IGHV"
                     /codon_start=3
                     /product="immunoglobulin heavy chain variable region"
                     /protein_id="CAD55894.1"
                     /db_xref="GI:24459871"
                     /translation="QAGNPGSASPRTTPRAKSLGSVSSVTTEDSATYYCTTVHQIRSY
                     GGYGYGGYGCYGYGYGYDYVDAWGQ"
     exon            <1. .209
                     /gene="IGHV"
                     /number=1
     V_region        <1. .>209
                     /gene="IGHV"
                     /product="immunoglobulin heavy chain variable region"
     V_segment       1. .108
                     /gene="IGHV"
     J_segment       109. .209
```

```
                     /gene="IGHV"
ORIGIN

Alignment Scores:
Pred. No.:                51.8          Length:         209
Score:                    42.00         Matches:        6
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              100.00%       Indels:         0
DB:                       4             Gaps:           0

US-09-858-332G-28 (1-6) x BTA512976 (1-209)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        186 CATAACCATAACCATAAC 169

RESULT 20
BTA512977/c
LOCUS       BTA512977               211 bp    DNA     linear   MAM 28-JUL-2004
DEFINITION  Bos taurus partial IGHV gene for immunoglobulin heavy chain
            variable region, exon 1 isolate fetus 1, clone 1.
ACCESSION   AJ512977
VERSION     AJ512977.1  GI:24459872
KEYWORDS    IGHV gene; immunoglobulin heavy chain; variable region.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1
  AUTHORS   Panelli,S., Damiani,G., Galli,C. and Sgaramella,V.
  TITLE     Rearranged genomes of bovine blood cells can allow the development
            of clones till late fetal stages; but rare unrearranged genomes
            have greater potential and lead to adulthood
  JOURNAL   Gene 334, 99-103 (2004)
   PUBMED   15256259
REFERENCE   2  (bases 1 to 211)
  AUTHORS   Panelli,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT-2002) Panelli S., Patol. Anim. Igiene e Sanita
            Pubbl. Vet., Fondazione PTP c/o Universita di Milano, via Celoria
            10, Milano 20133, ITALY
FEATURES             Location/Qualifiers
     source          1. .211
                     /organism="Bos taurus"
                     /mol_type="genomic DNA"
                     /isolate="fetus 1"
                     /db_xref="taxon:9913"
                     /clone="1"
                     /tissue_type="brain"
                     /dev_stage="fetus"
                     /rearranged
     gene            1. .211
                     /gene="IGHV"
     CDS             <1. .>211
                     /gene="IGHV"
                     /codon_start=2
                     /product="immunoglobulin heavy chain variable region"
                     /protein_id="CAD55895.1"
                     /db_xref="GI:24459873"
                     /translation="PGLKSRLSITKDNSKSQVSLSVSSVTTEDTATYYCSYCTTVHQI
                     SCPDGYSYGYGCGYGYGCSGYDCYGY"
     exon            <1. .211
                     /gene="IGHV"
                     /number=1
     V_region        <1. .>211
                     /gene="IGHV"
                     /product="immunoglobulin heavy chain variable region"
     V_segment       1. .114
                     /gene="IGHV"
     J_segment       115. .211
                     /gene="IGHV"

ORIGIN

Alignment Scores:
Pred. No.:                52.3          Length:         211
Score:                    42.00         Matches:        6
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              100.00%       Indels:         0
DB:                       4             Gaps:           0

US-09-858-332G-28 (1-6) x BTA512977 (1-211)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        182 CATAACCATAACCACAAC 165

RESULT 21
AY095308/c
LOCUS       AY095308                244 bp    DNA     linear   PLN 09-SEP-2003
DEFINITION  Amaranthus palmeri x Amaranthus rudis clone 21D1.244f hybrid
            progeny specific sequence.
ACCESSION   AY095308
VERSION     AY095308.1  GI:20977610
KEYWORDS    .
SOURCE      Amaranthus palmeri x Amaranthus rudis
  ORGANISM  Amaranthus palmeri x Amaranthus rudis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            Caryophyllales; Amaranthaceae; Amaranthus.
REFERENCE   1  (bases 1 to 244)
  AUTHORS   Steinau,A.N., Skinner,D.Z. and Steinau,M.
  TITLE     Mechanism of Extreme Genetic Recombination in Weedy Amaranthus
            Hybrids
  JOURNAL   Weed Sci. 51 (5), 696-701 (2003)
REFERENCE   2  (bases 1 to 244)
  AUTHORS   Steinau,A.N. and Skinner,D.Z.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2002) Crop and Soil Sciences, Kansas State
            University, Washington State University, 209 Johnson Hall, Pullman,
            WA 99164-6420, USA
FEATURES             Location/Qualifiers
     source          1. .244
                     /organism="Amaranthus palmeri x Amaranthus rudis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:194672"
                     /clone="21D1.244f"
     misc_feature    1. .244
                     /note="unique to interspecific hybrid progeny; similar to
                     Arabidopsis thaliana mutator transposase gene encoded by
                     GenBank Accession Number AB009049"
ORIGIN

Alignment Scores:
Pred. No.:                60.1          Length:         244
Score:                    42.00         Matches:        6
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              100.00%       Indels:         0
DB:                       15            Gaps:           0

US-09-858-332G-28 (1-6) x AY095308 (1-244)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        242 CACAATCACAATCACAAT 225

RESULT 22
AY095306/c
LOCUS       AY095306                263 bp    DNA     linear   PLN 09-SEP-2003
DEFINITION  Amaranthus palmeri x Amaranthus rudis clone 25d263f hybrid progeny
            specific sequence.
ACCESSION   AY095306
VERSION     AY095306.1  GI:20977608
```

```
KEYWORDS    .
SOURCE      Amaranthus palmeri x Amaranthus rudis
  ORGANISM  Amaranthus palmeri x Amaranthus rudis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            Caryophyllales; Amaranthaceae; Amaranthus.
REFERENCE   1  (bases 1 to 263)
  AUTHORS   Steinau,A.N., Skinner,D.Z. and Steinau,M.
  TITLE     Mechanism of Extreme Genetic Recombination in Weedy Amaranthus
            Hybrids
  JOURNAL   Weed Sci. 51 (5), 696-701 (2003)
REFERENCE   2  (bases 1 to 263)
  AUTHORS   Steinau,A.N. and Skinner,D.Z.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2002) Crop and Soil Sciences, Kansas State
            University, Washington State University, 209 Johnson Hall, Pullman,
            WA 99164-6420, USA
FEATURES             Location/Qualifiers
     source          1. .263
                     /organism="Amaranthus palmeri x Amaranthus rudis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:194672"
                     /clone="25d263f"
     misc_feature    1. .263
                     /note="unique to interspecific hybrid progeny; similar to
                     Arabidopsis thaliana mutator transposase gene encoded by
                     GenBank Accession Number AB009049"
ORIGIN

Alignment Scores:
Pred. No.:              64.6          Length:         263
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     15            Gaps:           0

US-09-858-332G-28 (1-6) x AY095306 (1-263)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        261 CACAATCACAATCACAAT 244

RESULT 23
AY095307/c
LOCUS       AY095307                267 bp    DNA     linear   PLN 09-SEP-2003
DEFINITION  Amaranthus palmeri x Amaranthus rudis clone 21a267f hybrid progeny
            specific sequence.
ACCESSION   AY095307
VERSION     AY095307.1  GI:20977609
KEYWORDS    .
SOURCE      Amaranthus palmeri x Amaranthus rudis
  ORGANISM  Amaranthus palmeri x Amaranthus rudis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            Caryophyllales; Amaranthaceae; Amaranthus.
REFERENCE   1  (bases 1 to 267)
  AUTHORS   Steinau,A.N., Skinner,D.Z. and Steinau,M.
  TITLE     Mechanism of Extreme Genetic Recombination in Weedy Amaranthus
            Hybrids
  JOURNAL   Weed Sci. 51 (5), 696-701 (2003)
REFERENCE   2  (bases 1 to 267)
  AUTHORS   Steinau,A.N. and Skinner,D.Z.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2002) Crop and Soil Sciences, Kansas State
            University, Washington State University, 209 Johnson Hall, Pullman,
            WA 99164-6420, USA
FEATURES             Location/Qualifiers
     source          1. .267
                     /organism="Amaranthus palmeri x Amaranthus rudis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:194672"
                     /clone="21a267f"
     misc_feature    1. .267
                     /note="unique to interspecific hybrid progeny; similar to
                     Arabidopsis thaliana mutator transposase gene encoded by
                     GenBank Accession Number AB009049"
ORIGIN

Alignment Scores:
Pred. No.:              65.6          Length:         267
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     15            Gaps:           0

US-09-858-332G-28 (1-6) x AY095307 (1-267)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        265 CACAATCACAATCACAAT 248

RESULT 24
SL12421/c
LOCUS       SL12421                 281 bp    DNA     linear   PLN 21-NOV-1997
DEFINITION  S.latifolia DNA containing repetitive element YG6.
ACCESSION   Y12421
VERSION     Y12421.1  GI:1934995
KEYWORDS    repetitive element.
SOURCE      Silene latifolia
  ORGANISM  Silene latifolia
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            Caryophyllales; Caryophyllaceae; Silene.
REFERENCE   1
  AUTHORS   Scutt,C.P., Kamisugi,Y., Sakai,F. and Gilmartin,P.M.
  TITLE     Laser isolation of plant sex chromosomes: studies on the DNA
            composition of the X and Y sex chromosomes of Silene latifolia
  JOURNAL   Genome 40 (5), 705-715 (1997)
   PUBMED   9352647
REFERENCE   2  (bases 1 to 281)
  AUTHORS   Scutt,C.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-1997) C.P. Scutt, University of Leeds, Centre for
            Plant Biochemistry and Biotechnology, Leeds, UK
FEATURES             Location/Qualifiers
     source          1. .281
                     /organism="Silene latifolia"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:37657"
                     /chromosome="Y"
                     /note="sequence found on all chromosomes"
     repeat_region   1. .281
                     /note="YG6 element"
ORIGIN

Alignment Scores:
Pred. No.:              68.9          Length:         281
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     15            Gaps:           0

US-09-858-332G-28 (1-6) x SL12421 (1-281)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        278 CACAACCACAATCACAAT 261

RESULT 25
AR014656/c
LOCUS       AR014656                292 bp    DNA     linear   PAT 05-DEC-1998
DEFINITION  Sequence 7 from patent US 5773693.
ACCESSION   AR014656
```

```
VERSION     AR014656.1  GI:3972110
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 292)
  AUTHORS   Burgess,D.G. and Dooner,H.K.
  TITLE     Pea ADP-glucose pyrophosphorylase subunit genes and their uses
  JOURNAL   Patent: US 5773693-A 7 30-JUN-1998;
FEATURES             Location/Qualifiers
     source          1. .292
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

Alignment Scores:
Pred. No.:                71.5          Length:          292
Score:                    42.00         Matches:         6
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       6             Gaps:            0

US-09-858-332G-28 (1-6) x AR014656 (1-292)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        206 CATAATCATAATCATAAT 189

RESULT 26
I18909/c
LOCUS       I18909                   292 bp    DNA     linear   PAT 07-OCT-1996
DEFINITION  Sequence 7 from patent US 5498831.
ACCESSION   I18909
VERSION     I18909.1  GI:1599264
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 292)
  AUTHORS   Burgess,D.G. and Dooner,H.K.
  TITLE     Pea ADP-glucose pyrophosphorylase subunit genes and their uses
  JOURNAL   Patent: US 5498831-A 7 12-MAR-1996;
FEATURES             Location/Qualifiers
     source          1. .292
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

Alignment Scores:
Pred. No.:                71.5          Length:          292
Score:                    42.00         Matches:         6
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       6             Gaps:            0

US-09-858-332G-28 (1-6) x I18909 (1-292)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        206 CATAATCATAATCATAAT 189

RESULT 27
BTU49780/c
LOCUS       BTU49780                 311 bp    mRNA    linear   MAM 16-AUG-2001
DEFINITION  Bos taurus immunoglobulin rearranged heavy chain variable region
            mRNA, partial cds.
ACCESSION   U49780
VERSION     U49780.1  GI:1293605
KEYWORDS    .
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 311)
  AUTHORS   Berens,S.J., Wylie,D.E. and Lopez,O.J.
  TITLE     Use of a single VH family and long CDR3s in the variable region of
            cattle Ig heavy chains
  JOURNAL   Int. Immunol. 9 (1), 189-199 (1997)
  PUBMED    9043960
REFERENCE   2  (bases 1 to 311)
  AUTHORS   Berens,S.J., Wylie,D. and Lopez,O.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-1996) Osvaldo J. Lopez, Biological Sciences,
            University of Nebraska, 325 Manter Hall, Lincoln, NE 68588-0118,
            USA
FEATURES             Location/Qualifiers
     source          1. .311
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /cell_type="heterohybridoma"
                     /tissue_type="spleen"
                     /dev_stage="150 day fetus"
     CDS             <1. .>311
                     /codon_start=3
                     /product="immunoglobulin heavy chain variable region"
                     /protein_id="AAA98661.1"
                     /db_xref="GI:1293606"
                     /translation="SGFSLSSYAVGWVRQAPGKALEWLGGISSGGSTYYNPALKSRLS
                     ITKDNSKSQVSLSVSSVTPEDTATYYCAKDGCYSGYGYGYGCGNYVDAWGQGLLVTVS
                     S"
ORIGIN

Alignment Scores:
Pred. No.:                76            Length:          311
Score:                    42.00         Matches:         6
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       4             Gaps:            0

US-09-858-332G-28 (1-6) x BTU49780 (1-311)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        261 CACAACCATAACCATAAC 244

RESULT 28
AJ748041/c
LOCUS       AJ748041                 317 bp    DNA     linear   INV 14-JUN-2004
DEFINITION  Trypanosoma cruzi kinetoplast minicircle, lineage IIe, strain CL
            Brener, isolate JP731.
ACCESSION   AJ748041
VERSION     AJ748041.1  GI:48714735
KEYWORDS    minicircle.
SOURCE      kinetoplast Trypanosoma cruzi
  ORGANISM  Trypanosoma cruzi
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma; Schizotrypanum.
REFERENCE   1
  AUTHORS   Telleria,J.
  TITLE     Analyse phylogenique intraspecifique de Trypanosoma cruzi par etude
            des sequences codantes exprimees et des sequences hypervariables
            des minicercles de l'ADN kinetoplastique. Implication sur le taux
            d'evolution clonal et de recombinaison genetique du parasite
  JOURNAL   Thesis (2003) Department of Ird, Universite Montpellier II,
            Montpellier, France
REFERENCE   2
  AUTHORS   Telleria,J., Lafay,B., Virreira,M., Barnabe,C., Tibayrenc,M. and
            Svoboda,M.
  TITLE     Sequence analysis of kinetoplastic minicercles DNA variability of
            Trypanosoma cruzi
  JOURNAL   Unpublished
```

```
REFERENCE   3  (bases 1 to 317)
  AUTHORS   Svoboda,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAY-2004) Svoboda M., Chimie Biologique, Faculte de
            Medecine, Route de Lennik 808, CP 611, 1070 Bruxelles, BELGIUM
FEATURES             Location/Qualifiers
     source          1. .317
                     /organism="Trypanosoma cruzi"
                     /organelle="mitochondrion:kinetoplast"
                     /mol_type="genomic DNA"
                     /strain="CL Brener"
                     /isolate="JP731"
                     /specific_host="Dasyprocta aguti"
                     /db_xref="taxon:5693"
                     /country="Brazil:Espiruto Santo"
                     /note="lineage IIe"
     primer_bind     1. .26
     misc_feature    27. .275
                     /note="variable domain"
     misc_feature    276. .290
                     /note="constant domain"
     primer_bind     complement(292. .317)
ORIGIN

Alignment Scores:
Pred. No.:              77.4          Length:         317
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     2             Gaps:           0

US-09-858-332G-28 (1-6) x AJ748041 (1-317)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        203 CACAACCACAACCACAAC 186

RESULT 29
AR552556
LOCUS       AR552556                339 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 7687 from patent US 6747137.
ACCESSION   AR552556
VERSION     AR552556.1  GI:53945731
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 339)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 7687 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .339
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              82.5          Length:         339
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     6             Gaps:           0

US-09-858-332G-28 (1-6) x AR552556 (1-339)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        200 CACAACCACAACCACAAC 217

RESULT 30
AF058630/c
LOCUS       AF058630                346 bp    DNA     linear   MAM 01-DEC-1998
DEFINITION  Oryctolagus cuniculus clone Mb_60-H immunoglobulin heavy chain
            V-D-J region (IgH) gene, partial cds.
ACCESSION   AF058630
VERSION     AF058630.1  GI:3930461
KEYWORDS    .
SOURCE      Oryctolagus cuniculus (rabbit)
  ORGANISM  Oryctolagus cuniculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Lagomorpha;
            Leporidae; Oryctolagus.
REFERENCE   1  (bases 1 to 346)
  AUTHORS   Sehgal,D., Schiaffella,E., Anderson,A.O. and Mage,R.G.
  TITLE     Analyses of single B cells by polymerase chain reaction reveal
            rearranged VH with germline sequences in spleens of immunized adult
            rabbits: implications for B cell repertoire maintenance and renewal
  JOURNAL   J. Immunol. 161 (10), 5347-5356 (1998)
   PUBMED   9820508
REFERENCE   2  (bases 1 to 346)
  AUTHORS   Mage,R.G., Sehgal,D., Schiaffella,E. and Anderson,A.O.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-APR-1998) Laboratory of Immunology, National
            Institute of Allergy and Infectious Diseases, Bldg. 10; Rm #11N311,
            10 Center Drive - MSC 1892, National Institutes of Health,
            Bethesda, MD 20892-1892, USA
FEATURES             Location/Qualifiers
     source          1. .346
                     /organism="Oryctolagus cuniculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9986"
                     /clone="Mb_60-H"
     gene            <1. .>346
                     /gene="IgH"
     CDS             <1. .>346
                     /gene="IgH"
                     /codon_start=1
                     /product="immunoglobulin heavy chain V-D-J region"
                     /protein_id="AAC80399.1"
                     /db_xref="GI:3930462"
                     /translation="QSVKESEGGLFKPTDTLTLTCTVSGFSLSTYGVNWVRQAPGKGL
                     EWIGGIGSNGDLYYASWAKSRSTITRNTNLNTVTLKMTSLTAADTATYFCAREGGYGY
                     GYAFSIWGPGTLV"
ORIGIN

Alignment Scores:
Pred. No.:              84.2          Length:         346
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     4             Gaps:           0

US-09-858-332G-28 (1-6) x AF058630 (1-346)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        313 CATAACCATAACCATAAC 296

RESULT 31
AF434459/c
LOCUS       AF434459                346 bp    DNA     linear   MAM 30-MAY-2002
DEFINITION  Oryctolagus cuniculus RUa-061-H nonfunctional immunoglobulin heavy
            chain V-D-J region (IgH) gene, partial sequence.
ACCESSION   AF434459
VERSION     AF434459.1  GI:21262573
KEYWORDS    .
SOURCE      Oryctolagus cuniculus (rabbit)
  ORGANISM  Oryctolagus cuniculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Lagomorpha;
```

```
            Leporidae; Oryctolagus.
REFERENCE   1  (bases 1 to 346)
  AUTHORS   Sehgal,D., Obiakor,H. and Mage,R.G.
  TITLE     Distinct clonal Ig diversification patterns in young appendix
            compared to antigen-specific splenic clones
  JOURNAL   J. Immunol. 168 (11), 5424-5433 (2002)
   PUBMED   12023335
REFERENCE   2  (bases 1 to 346)
  AUTHORS   Mage,R.G., Sehgal,D. and Obiakor,H.T.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-OCT-2001) Laboratory of Immunology, NIAID, National
            Institutes of Health, 10 Center Drive - MSC - 1892, Bethesda, MD
            20892-1892, USA
FEATURES             Location/Qualifiers
     source          1. .346
                     /organism="Oryctolagus cuniculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9986"
                     /clone="RUa-061-H"
                     /haplotype="F-I a2/a2 b5/b5"
                     /cell_type="B-cell"
                     /tissue_type="appendix"
                     /note="sequence obtained from germinal center B-cell by
                     single cell PCR"
     gene            <1. .>346
                     /gene="IgH"
     misc_feature    <1. .>346
                     /gene="IgH"
                     /note="nonfunctional immunoglobulin heavy chain V-D-J
                     region"
ORIGIN

Alignment Scores:
Pred. No.:                 84.2          Length:          346
Score:                     42.00         Matches:         6
Percent Similarity:        100.00%       Conservative:    0
Best Local Similarity:     100.00%       Mismatches:      0
Query Match:               100.00%       Indels:          0
DB:                        4             Gaps:            0

US-09-858-332G-28 (1-6) x AF434459 (1-346)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        316 CATAACCATAACCATAAC 299

RESULT 32
AF434364/c
LOCUS       AF434364                 349 bp    DNA     linear   MAM 30-MAY-2002
DEFINITION  Oryctolagus cuniculus RLa-021-H immunoglobulin heavy chain V-D-J
            region (IgH) gene, partial cds.
ACCESSION   AF434364
VERSION     AF434364.1  GI:21262392
KEYWORDS    .
SOURCE      Oryctolagus cuniculus (rabbit)
  ORGANISM  Oryctolagus cuniculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Lagomorpha;
            Leporidae; Oryctolagus.
REFERENCE   1  (bases 1 to 349)
  AUTHORS   Sehgal,D., Obiakor,H. and Mage,R.G.
  TITLE     Distinct clonal Ig diversification patterns in young appendix
            compared to antigen-specific splenic clones
  JOURNAL   J. Immunol. 168 (11), 5424-5433 (2002)
   PUBMED   12023335
REFERENCE   2  (bases 1 to 349)
  AUTHORS   Mage,R.G., Sehgal,D. and Obiakor,H.T.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-OCT-2001) Laboratory of Immunology, NIAID, National
            Institutes of Health, 10 Center Drive - MSC - 1892, Bethesda, MD
            20892-1892, USA
FEATURES             Location/Qualifiers
     source          1. .349
                     /organism="Oryctolagus cuniculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9986"
                     /clone="RLa-021-H"
                     /haplotype="F-I a2/a2 b5/b5"
                     /cell_type="B-cell"
                     /tissue_type="appendix"
                     /note="sequence obtained from germinal center B-cell by
                     single cell PCR"
     gene            <1. .>349
                     /gene="IgH"
     CDS             <1. .>349
                     /gene="IgH"
                     /codon_start=1
                     /product="immunoglobulin heavy chain V-D-J region"
                     /protein_id="AAM44546.1"
                     /db_xref="GI:21262393"
                     /translation="QSVKESEGGLFKPTDTLTLTCTVSGFSLSSYGVSWVRQAPGNGL
                     EWIGAIGSSGSAYYASWAKSRSTITRNTNLNTVTLKMTSLTAADTATYFCARDRGSYG
                     YDYGKDIWGPGTLV"
ORIGIN

Alignment Scores:
Pred. No.:                 84.9          Length:          349
Score:                     42.00         Matches:         6
Percent Similarity:        100.00%       Conservative:    0
Best Local Similarity:     100.00%       Mismatches:      0
Query Match:               100.00%       Indels:          0
DB:                        4             Gaps:            0

US-09-858-332G-28 (1-6) x AF434364 (1-349)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        316 CATAATCATAACCATAAC 299

RESULT 33
AF097208/c
LOCUS       AF097208                 351 bp    mRNA    linear   MAM 17-DEC-2003
DEFINITION  Bos taurus clone BF3A11 immunoglobulin heavy chain variable region
            mRNA, partial cds.
ACCESSION   AF097208
VERSION     AF097208.1  GI:3834656
KEYWORDS    .
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 351)
  AUTHORS   Saini,S.S. and Kaushik,A.
  TITLE     Extensive CDR3H length heterogeneity exists in bovine foetal VDJ
            rearrangements
  JOURNAL   Scand. J. Immunol. 55 (2), 140-148 (2002)
   PUBMED   11896930
REFERENCE   2  (bases 1 to 351)
  AUTHORS   Saini,S.S., Allore,B., Jacobs,R. and Kaushik,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1998) Pathobiology, University of Guelph, Guelph,
            ON N1G 2W1, Canada
FEATURES             Location/Qualifiers
     source          1. .351
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /clone="BF3A11"
                     /cell_type="mouse x Bovine heterohybridoma"
                     /dev_stage="fetal"
                     /note="PCR product"
     CDS             <1. .>351
                     /note="VDJ region"
                     /codon_start=1
                     /product="immunoglobulin heavy chain variable region"
```

```
                     /protein_id="AAC71043.1"
                     /db_xref="GI:3834657"
                     /translation="QVQLRESGPSLVKPSQTLSLTCTVSGFSLSDNSVGWVRQAPGKA
                     LEWLGVIYSGGSTGYNPALKSRLSITKDNSKSQVSLSLSSVTTEDTATYYCARGLVMV
                     MVVTWGQGLLVTVSS"
ORIGIN

Alignment Scores:
Pred. No.:                 85.4          Length:          351
Score:                     42.00         Matches:         6
Percent Similarity:        100.00%       Conservative:    0
Best Local Similarity:     100.00%       Mismatches:      0
Query Match:               100.00%       Indels:          0
DB:                        4             Gaps:            0

US-09-858-332G-28 (1-6) x AF097208 (1-351)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        315 CACAACCATAACCATAAC 298

RESULT 34
AY225843
LOCUS       AY225843                 355 bp    DNA     linear   PLN 01-MAY-2003
DEFINITION  Amicia glandulosa isolate 1 LEGCYC gene, partial cds.
ACCESSION   AY225843
VERSION     AY225843.1  GI:30313970
KEYWORDS    .
SOURCE      Amicia glandulosa
  ORGANISM  Amicia glandulosa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae;
            Aeschynomeneae; Amicia.
REFERENCE   1  (bases 1 to 355)
  AUTHORS   Citerne,H.L., Luo,D., Pennington,R.T., Coen,E. and Cronk,Q.C.
  TITLE     A Phylogenomic Investigation of CYCLOIDEA-Like TCP Genes in the
            Leguminosae
  JOURNAL   Plant Physiol. 131 (3), 1042-1053 (2003)
   PUBMED   12644657
REFERENCE   2  (bases 1 to 355)
  AUTHORS   Citerne,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) Royal Botanic Garden Edinburgh, 20a
            Inverleith Row, Edinburgh EH3 5LR, U.K.
FEATURES             Location/Qualifiers
     source          1. .355
                     /organism="Amicia glandulosa"
                     /mol_type="genomic DNA"
                     /isolate="1"
                     /specimen_voucher="R.T. Pennington 654"
                     /db_xref="taxon:105934"
                     /note="authority: Amicia glandulosa Kunth"
     mRNA            <1. .>355
                     /product="LEGCYC"
     CDS             <1. .>355
                     /codon_start=2
                     /product="LEGCYC"
                     /protein_id="AAO88045.1"
                     /db_xref="GI:30313971"
                     /translation="RVRLSIEIARKFFDLQDMSGFDKASNTLEWLLNKSKKAIKDLAR
                     TTNNHNHNHNHNSSFSNYNSSSSDEGEVDSVINHQQDHEEEEEEEEAGPNGNSEELKG
                     KKMMMRREKMKESREK"
ORIGIN

Alignment Scores:
Pred. No.:                 86.3          Length:          355
Score:                     42.00         Matches:         6
Percent Similarity:        100.00%       Conservative:    0
Best Local Similarity:     100.00%       Mismatches:      0
Query Match:               100.00%       Indels:          0
DB:                        15            Gaps:            0

US-09-858-332G-28 (1-6) x AY225843 (1-355)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        146 CACAACCACAACCACAAT 163

RESULT 35
BTU49777/c
LOCUS       BTU49777                 360 bp    mRNA    linear   MAM 16-AUG-2001
DEFINITION  Bos taurus immunoglobulin rearranged heavy chain variable region
            mRNA, partial cds.
ACCESSION   U49777
VERSION     U49777.1  GI:1293599
KEYWORDS    .
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 360)
  AUTHORS   Berens,S.J., Wylie,D.E. and Lopez,O.J.
  TITLE     Use of a single VH family and long CDR3s in the variable region of
            cattle Ig heavy chains
  JOURNAL   Int. Immunol. 9 (1), 189-199 (1997)
   PUBMED   9043960
REFERENCE   2  (bases 1 to 360)
  AUTHORS   Berens,S.J., Wylie,D. and Lopez,O.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-1996) Osvaldo J. Lopez, Biological Sciences,
            University of Nebraska, 325 Manter Hall, Lincoln, NE 68588-0118,
            USA
FEATURES             Location/Qualifiers
     source          1. .360
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /cell_type="heterohybridoma"
                     /tissue_type="spleen"
                     /dev_stage="150 day fetus"
     CDS             <1. .>360
                     /codon_start=1
                     /product="immunoglobulin heavy chain variable region"
                     /protein_id="AAA98658.1"
                     /db_xref="GI:1293600"
                     /translation="LVKPSQTLSLTCTVSGFSLSSYGVGWVRQAPGKALECLGGISSG
                     GSTGYNPALKSRLSITKDNSKSQVSLSLSSVTTEDTATYYCAKYGGYGYGGYGCYGYG
                     YGYGIDAWGQGLLVTVSS"
ORIGIN

Alignment Scores:
Pred. No.:                 87.5           Length:          360
Score:                     42.00          Matches:         6
Percent Similarity:        100.00%        Conservative:    0
Best Local Similarity:     100.00%        Mismatches:      0
Query Match:               100.00%        Indels:          0
DB:                        4              Gaps:            0

US-09-858-332G-28 (1-6) x BTU49777 (1-360)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        316 CATAACCATAACCATAAC 299

RESULT 36
AF058641/c
LOCUS       AF058641                 364 bp    DNA     linear   MAM 01-DEC-1998
DEFINITION  Oryctolagus cuniculus clone Mb_02-H immunoglobulin heavy chain
            V-D-J region (IgH) gene, partial cds.
ACCESSION   AF058641
VERSION     AF058641.1  GI:3930483
KEYWORDS    .
SOURCE      Oryctolagus cuniculus (rabbit)
  ORGANISM  Oryctolagus cuniculus
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Lagomorpha; Leporidae; Oryctolagus.
REFERENCE 1 (bases 1 to 364)
AUTHORS Sehgal,D., Schiaffella,E., Anderson,A.O. and Mage,R.G.
TITLE Analyses of single B cells by polymerase chain reaction reveal rearranged VH with germline sequences in spleens of immunized adult rabbits: implications for B cell repertoire maintenance and renewal
JOURNAL J. Immunol. 161 (10), 5347-5356 (1998)
PUBMED 9820508
REFERENCE 2 (bases 1 to 364)
AUTHORS Mage,R.G., Sehgal,D., Schiaffella,E. and Anderson,A.O.
TITLE Direct Submission
JOURNAL Submitted (09-APR-1998) Laboratory of Immunology, National Institute of Allergy and Infectious Diseases, Bldg. 10; Rm #11N311, 10 Center Drive - MSC 1892, National Institutes of Health, Bethesda, MD 20892-1892, USA

```
FEATURES             Location/Qualifiers
     source          1. .364
                     /organism="Oryctolagus cuniculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9986"
                     /clone="Mb_02-H"
     gene            <1. .>364
                     /gene="IgH"
     CDS             <1. .>364
                     /gene="IgH"
                     /codon_start=1
                     /product="immunoglobulin heavy chain V-D-J region"
                     /protein_id="AAC80410.1"
                     /db_xref="GI:3930484"
                     /translation="QSVKESEGGLFKPTDTLKLTCTVSGFSLSYNAISWVRQAPGNGL
                     EWIGAIGSSGSAYYASWAKSRSTITRNTNLNTVTLKMTSLTAADTATYFCAREASYGY
                     GYGIGRSDAFDPWGPGTLV"
ORIGIN
```

Alignment Scores:

Pred. No.:	88.4	Length:	364
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	4	Gaps:	0

US-09-858-332G-28 (1-6) x AF058641 (1-364)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        313 CATAACCATAACCATAAC 296
```

RESULT 37
AJ836260/c
LOCUS AJ836260 366 bp DNA linear PLN 22-SEP-2004
DEFINITION Arabidopsis thaliana T-DNA flanking sequence, left border, clone 201C09.
ACCESSION AJ836260
VERSION AJ836260.1 GI:52540466
KEYWORDS left border; T-DNA flanking sequence.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Brunaud,V., Balzergue,S., Dubreucq,B., Aubourg,S., Samson,F., Chauvin,S., Bechtold,N., Cruaud,C., DeRose,R., Pelletier,G., Lepiniec,L., Caboche,M. and Lecharny,A.
TITLE T-DNA integration into the Arabidopsis genome depends on sequences of pre-insertion sites
JOURNAL EMBO Rep. 3 (12), 1152-1157 (2002)
PUBMED 12446565
REFERENCE 2 (bases 1 to 366)
AUTHORS Balzergue,S.
TITLE Direct Submission
JOURNAL Submitted (21-SEP-2004) Balzergue S., UMRGV, INRA/CNRS, 2 rue Gaston Cremieux, 91057 Evry cedex, FRANCE
COMMENT PCR was performed on DNA from transformants of Arabidopsis thaliana plants from INRA (Versailles). The DNA fragment(s) resulting from the PCR were directly sequenced from the left or the right border to determine the genomic sequence flanking the insertion. T-DNA derived sequences were removed. Information to order the corresponding mutant line and a link to a database providing a graphical display of the insertion site are available at http://dbsgap.versailles.inra.fr/publiclines/. This sequence has been generated in the framework of the French plant genomics program 'Genoplante' (http://www.genoplante.com and http://genoplante-info.infobiogen.fr).

```
FEATURES             Location/Qualifiers
     source          1. .366
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /clone="201C09"
                     /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
                     /ecotype="Wassilewskija"
     misc_feature    1. .366
                     /note="T-DNA flanking sequence
                     left border"
ORIGIN
```

Alignment Scores:

Pred. No.:	88.9	Length:	366
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	15	Gaps:	0

US-09-858-332G-28 (1-6) x AJ836260 (1-366)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        291 CACAATCACAATCATAAT 274
```

RESULT 38
MMU40980/c
LOCUS MMU40980 372 bp mRNA linear ROD 08-AUG-1996
DEFINITION Mus musculus trp-related protein 1 mRNA, partial cds.
ACCESSION U40980
VERSION U40980.1 GI:1373058
KEYWORDS .
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 372)
AUTHORS Zhu,X., Jiang,M., Peyton,M., Boulay,G., Hurst,R., Stefani,E. and Birnbaumer,L.
TITLE trp, a novel mammalian gene family essential for agonist-activated capacitative Ca2+ entry
JOURNAL Cell 85 (5), 661-671 (1996)
PUBMED 8646775
REFERENCE 2 (bases 1 to 372)
AUTHORS Zhu,X.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-1995) Xi Zhu, Anesthesiology, UCLA, 10833 Le Conte Ave., BH612, CHS, Los Angeles, CA 90095-1778, USA

```
FEATURES             Location/Qualifiers
     source          1. .372
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
     CDS             <1. .>372
                     /note="Drosophila trp homolog"
                     /codon_start=1
```

```
                     /product="trp-related protein 1"
                     /protein_id="AAC52699.1"
                     /db_xref="GI:1373059"
                     /translation="QDFGKFLGMFLLVLFSFTIGLTQLYDKGYTSKEQKDCVGIFCEQ
                     QSNDTFHSFIGTCFALFWYIFSLAHVAIFVTRFSYGEELQSFVGAVIVGTYNVVVVIV
                     LTKLLVAMLHKSFQLIANHEDK"
ORIGIN

Alignment Scores:
Pred. No.:                90.3          Length:        372
Score:                    42.00         Matches:       6
Percent Similarity:       100.00%       Conservative:  0
Best Local Similarity:    100.00%       Mismatches:    0
Query Match:              100.00%       Indels:        0
DB:                       9             Gaps:          0

US-09-858-332G-28 (1-6) x MMU40980 (1-372)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        306 CACAATCACAACCACAAC 289

RESULT 39
AF097206/c
LOCUS       AF097206                381 bp    mRNA    linear   MAM 17-DEC-2003
DEFINITION  Bos taurus clone BF2D12 immunoglobulin heavy chain variable region
            mRNA, partial cds.
ACCESSION   AF097206
VERSION     AF097206.1  GI:3834652
KEYWORDS    .
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 381)
  AUTHORS   Saini,S.S. and Kaushik,A.
  TITLE     Extensive CDR3H length heterogeneity exists in bovine foetal VDJ
            rearrangements
  JOURNAL   Scand. J. Immunol. 55 (2), 140-148 (2002)
   PUBMED   11896930
REFERENCE   2  (bases 1 to 381)
  AUTHORS   Saini,S.S., Allore,B., Jacobs,R. and Kaushik,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1998) Pathobiology, University of Guelph, Guelph,
            ON N1G 2W1, Canada
FEATURES             Location/Qualifiers
     source          1. .381
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /clone="BF2D12"
                     /cell_type="mouse x Bovine heterohybridoma"
                     /dev_stage="fetal"
                     /note="PCR product"
     CDS             <1. .>381
                     /note="VDJ region"
                     /codon_start=1
                     /product="immunoglobulin heavy chain variable region"
                     /protein_id="AAC71041.1"
                     /db_xref="GI:3834653"
                     /translation="QVQLRESGPSLVKPSQTLSLTCTISGFSLSSYAVGWVRQAPGKA
                     LEWLGGISSGGSTCLNPALKSRLSITKDNSKSQVSLSVSSVTTEDTATYYCTCYSGYG
                     YGYGYAYGYDYVDAWGQGLLVTVSS"
ORIGIN

Alignment Scores:
Pred. No.:                92.4          Length:        381
Score:                    42.00         Matches:       6
Percent Similarity:       100.00%       Conservative:  0
Best Local Similarity:    100.00%       Mismatches:    0
Query Match:              100.00%       Indels:        0
DB:                       4             Gaps:          0

US-09-858-332G-28 (1-6) x AF097206 (1-381)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        322 CATAACCATAACCATAAC 305

RESULT 40
AF097207/c
LOCUS       AF097207                387 bp    mRNA    linear   MAM 17-DEC-2003
DEFINITION  Bos taurus clone BF3H11 immunoglobulin heavy chain variable region
            mRNA, partial cds.
ACCESSION   AF097207
VERSION     AF097207.1  GI:3834654
KEYWORDS    .
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 387)
  AUTHORS   Saini,S.S. and Kaushik,A.
  TITLE     Extensive CDR3H length heterogeneity exists in bovine foetal VDJ
            rearrangements
  JOURNAL   Scand. J. Immunol. 55 (2), 140-148 (2002)
   PUBMED   11896930
REFERENCE   2  (bases 1 to 387)
  AUTHORS   Saini,S.S., Allore,B., Jacobs,R. and Kaushik,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1998) Pathobiology, University of Guelph, Guelph,
            ON N1G 2W1, Canada
FEATURES             Location/Qualifiers
     source          1. .387
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /clone="BF3H11"
                     /cell_type="mouse x Bovine heterohybridoma"
                     /dev_stage="fetal"
                     /note="PCR product"
     CDS             <1. .>387
                     /note="VDJ region"
                     /codon_start=1
                     /product="immunoglobulin heavy chain variable region"
                     /protein_id="AAC71042.1"
                     /db_xref="GI:3834655"
                     /translation="QVQLRESGPSLVKPSQTLSLTCTVSGFSLSSNGVGWVRQAPGKA
                     LEWVGGIDNDGDTYYNPALKSRLSITKDNSKSQVSLSVSSVTPEDTATYYCANSCYSG
                     YGYGCGYGYGYDYVDAWGQGLLVTVSS"
ORIGIN

Alignment Scores:
Pred. No.:                93.8          Length:        387
Score:                    42.00         Matches:       6
Percent Similarity:       100.00%       Conservative:  0
Best Local Similarity:    100.00%       Mismatches:    0
Query Match:              100.00%       Indels:        0
DB:                       4             Gaps:          0

US-09-858-332G-28 (1-6) x AF097207 (1-387)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        340 CATAACCATAACCATAAC 323

RESULT 41
AF097205/c
LOCUS       AF097205                390 bp    mRNA    linear   MAM 17-DEC-2003
DEFINITION  Bos taurus clone BF2B5 immunoglobulin heavy chain variable region
            mRNA, partial cds.
ACCESSION   AF097205
VERSION     AF097205.1  GI:3834650
KEYWORDS    .
```

```
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 390)
  AUTHORS   Saini,S.S. and Kaushik,A.
  TITLE     Extensive CDR3H length heterogeneity exists in bovine foetal VDJ
            rearrangements
  JOURNAL   Scand. J. Immunol. 55 (2), 140-148 (2002)
   PUBMED   11896930
REFERENCE   2  (bases 1 to 390)
  AUTHORS   Saini,S.S., Allore,B., Jacobs,R. and Kaushik,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1998) Pathobiology, University of Guelph, Guelph,
            ON N1G 2W1, Canada
FEATURES             Location/Qualifiers
     source          1. .390
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /clone="BF2B5"
                     /cell_type="mouse x Bovine heterohybridoma"
                     /dev_stage="fetal"
                     /note="PCR product"
     CDS             <1. .>390
                     /note="VDJ region"
                     /codon_start=1
                     /product="immunoglobulin heavy chain variable region"
                     /protein_id="AAC71040.1"
                     /db_xref="GI:3834651"
                     /translation="QVQLRESGPSLVKPSQTLSLTCTVFGFSLSDNSVGWVRQVPGKA
                     LEWLGVIYSGGSTGYNPALKSRLSITKDNSKSQVSLSLSSVTTEDTATYYCARDCYSD
                     YGNGYGCGYGYGSYVDAWGQGLLVTVSS"
ORIGIN

Alignment Scores:
Pred. No.:                94.5          Length:          390
Score:                    42.00         Matches:         6
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       4             Gaps:            0

US-09-858-332G-28 (1-6) x AF097205 (1-390)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        340 CATAACCATAACCACAAC 323

RESULT 42
AF097209/c
LOCUS       AF097209                390 bp    mRNA    linear   MAM 17-DEC-2003
DEFINITION  Bos taurus clone BF5F10 immunoglobulin heavy chain variable region
            mRNA, partial cds.
ACCESSION   AF097209
VERSION     AF097209.1  GI:3834658
KEYWORDS    .
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 390)
  AUTHORS   Saini,S.S. and Kaushik,A.
  TITLE     Extensive CDR3H length heterogeneity exists in bovine foetal VDJ
            rearrangements
  JOURNAL   Scand. J. Immunol. 55 (2), 140-148 (2002)
   PUBMED   11896930
REFERENCE   2  (bases 1 to 390)
  AUTHORS   Saini,S.S., Allore,B., Jacobs,R. and Kaushik,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1998) Pathobiology, University of Guelph, Guelph,
            ON N1G 2W1, Canada
FEATURES             Location/Qualifiers
     source          1. .390
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /clone="BF5F10"
                     /cell_type="mouse x Bovine heterohybridoma"
                     /dev_stage="fetal"
                     /note="PCR product"
     CDS             <1. .>390
                     /note="VDJ region"
                     /codon_start=1
                     /product="immunoglobulin heavy chain variable region"
                     /protein_id="AAC71044.1"
                     /db_xref="GI:3834659"
                     /translation="QVQLRESGPSLVKPSQTLSLTCTVSGFSLSDNSVGWVRQAPGKA
                     LEWLGVIYSGGSTGYNPALKSRLSITKDNSKSQVSLSLSSVTTEDTATYYCARDCYSG
                     YGYGYGCGYGYGSYVDAWGQGLLVTVSS"
ORIGIN

Alignment Scores:
Pred. No.:                94.5          Length:          390
Score:                    42.00         Matches:         6
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       4             Gaps:            0

US-09-858-332G-28 (1-6) x AF097209 (1-390)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        340 CATAACCATAACCACAAC 323

RESULT 43
PFA430086
LOCUS       PFA430086               397 bp    DNA     linear   INV 07-FEB-2002
DEFINITION  Plasmodium falciparum partial rh1 gene, isolate 3D7.
ACCESSION   AJ430086
VERSION     AJ430086.1  GI:18642427
KEYWORDS    rh1 gene.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1
  AUTHORS   Taylor,H.M., Grainger,M. and Holder,A.A.
  TITLE     Variation in the expression of a Plasmodium falciparum protein
            family implicated in erythrocyte invasion
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 397)
  AUTHORS   Taylor,H.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-FEB-2002) Taylor H.M., Division of Parasitology,
            National Institute for Medical Research, Mill Hill, London, NW7
            1AA, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1. .397
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="3D7"
                     /db_xref="taxon:5833"
     gene            1. .396
                     /gene="rh1"
     CDS             <1. .>396
                     /gene="rh1"
                     /function="putative role in erythrocyte invasion"
                     /codon_start=1
                     /product="RH1 protein"
                     /protein_id="CAD23026.1"
                     /db_xref="GI:18642428"
                     /db_xref="UniProt/TrEMBL:Q8T8F3"
                     /translation="LENIKNNENIFDNIQLKKKDIDDIIININNTKETYLKELNKKKM
```

LQNKKKVDEKSEINNHHTLQHDNQNVEQKNKIKDHNLITKPNNNSSEESHQNEQMKEQ
NKNILEKQTRNIKPHHVHNHNHNHNQNQKD"
ORIGIN

Alignment Scores:

Pred. No.:	96.1	Length:	397
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-28 (1-6) x PFA430086 (1-397)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        358 CATAATCATAATCATAAT 375
```

RESULT 44
BTU49781/c
LOCUS BTU49781 408 bp mRNA linear MAM 16-AUG-2001
DEFINITION Bos taurus immunoglobulin rearranged heavy chain variable region mRNA, partial cds.
ACCESSION U49781
VERSION U49781.1 GI:1293607
KEYWORDS .
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 408)
AUTHORS Berens,S.J., Wylie,D.E. and Lopez,O.J.
TITLE Use of a single VH family and long CDR3s in the variable region of cattle Ig heavy chains
JOURNAL Int. Immunol. 9 (1), 189-199 (1997)
PUBMED 9043960
REFERENCE 2 (bases 1 to 408)
AUTHORS Berens,S.J., Wylie,D. and Lopez,O.J.
TITLE Direct Submission
JOURNAL Submitted (22-FEB-1996) Osvaldo J. Lopez, Biological Sciences, University of Nebraska, 325 Manter Hall, Lincoln, NE 68588-0118, USA
FEATURES Location/Qualifiers
source 1. .408
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/cell_type="heterohybridoma"
/tissue_type="spleen"
/dev_stage="150 day fetus"
CDS <1. .>408
/codon_start=1
/product="immunoglobulin heavy chain variable region"
/protein_id="AAA98662.1"
/db_xref="GI:1293608"
/translation="FGLFTLSDKAVGWVRQAPGKALEWLGGIDTGGSTGYNPGLKSRL
SITKDNSKSQVSLSVSSVTTEDSATYYCTTVHQIRCPDGYGYGYGCGYGSYGYSGYDC
YGYGGYGGYGGYGGYSSYSYVDAWGQGLLVTVSS"
ORIGIN

Alignment Scores:

Pred. No.:	98.7	Length:	408
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	4	Gaps:	0

US-09-858-332G-28 (1-6) x BTU49781 (1-408)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        277 CATAACCACAACCATAAC 260
```

RESULT 45
PFA430087
LOCUS PFA430087 409 bp DNA linear INV 07-FEB-2002
DEFINITION Plasmodium falciparum partial rh1 gene, isolate T996.
ACCESSION AJ430087
VERSION AJ430087.1 GI:18642429
KEYWORDS rh1 gene.
SOURCE Plasmodium falciparum (malaria parasite P. falciparum)
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1
AUTHORS Taylor,H.M., Grainger,M. and Holder,A.A.
TITLE Variation in the expression of a Plasmodium falciparum protein family implicated in erythrocyte invasion
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 409)
AUTHORS Taylor,H.M.
TITLE Direct Submission
JOURNAL Submitted (04-FEB-2002) Taylor H.M., Division of Parasitology, National Institute for Medical Research, Mill Hill, London, NW7 1AA, UNITED KINGDOM
FEATURES Location/Qualifiers
source 1. .409
/organism="Plasmodium falciparum"
/mol_type="genomic DNA"
/isolate="T996"
/db_xref="taxon:5833"
/country="Thailand"
gene 1. .408
/gene="rh1"
CDS <1. .>408
/gene="rh1"
/function="putative role in erythrocyte invasion"
/codon_start=1
/product="RH1 protein"
/protein_id="CAD23027.1"
/db_xref="GI:18642430"
/db_xref="UniProt/TrEMBL:Q8T8F2"
/translation="LENIKNNENIFDNIQLKKKDIDDIIININNTKETYLKELNKKKM
LQNKKKVDEKSEINNHHTLQHDNQNVEQKNKIKDHNLITKPNNNSSEESHQNEQMKEQ
NKNILEKQTRNIKPHHVHNHNHNHNQNQNQNQKD"
ORIGIN

Alignment Scores:

Pred. No.:	98.9	Length:	409
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-28 (1-6) x PFA430087 (1-409)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        358 CATAATCATAATCATAAT 375
```

RESULT 46
G46669
LOCUS G46669 425 bp DNA linear STS 23-MAR-1999
DEFINITION Z14484_1 Zebrafish AB Danio rerio STS genomic clone Z14484 5', sequence tagged site.
ACCESSION G46669
VERSION G46669.1 GI:4492960
KEYWORDS STS.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 425)

```
  AUTHORS   Ziniti,J., Jackson,D.G. and Fishman,M.C.
  TITLE     A Zebrafish microsatellite map
  JOURNAL   Unpublished (1999)
COMMENT
            Contact: Mark C. Fishman
            Cardiovascular Research Center
            Massachusetts General Hospital
            Mail code 1494100A, 149 13th Street, Charlestown, MA 02129, USA
            Fax: 6177265806
            Email: fishman@mgh.cvrc.harvard.edu
            http://zebrafish.mgh.harvard.edu
            Primer A: TGTGTGTGTGTGCTTGCATT
            Primer B: AGTTGTCAGCGTTCTGGCTT
            STS size: 152
            PCR Profile:
                Presoak:        94 degrees C for 5.0 minutes
                Denaturation:   94 degrees C for 1.0 minute
                Annealing:      58 degrees C for 1.0 minute
                Polymerization: 72 degrees C for 1.5 minute
                PCR Cycles:     27
                Thermal Cycler: MJ Research PTC-100
            Protocol:
                Template:       10 ng
                Primer:         each 375 nM
                dNTPs:          each 200 uM
                Taq Polymerase: 0.034 units/ul
                Total Vol:      10 ul

            Buffer:
                            MgCL2:          1.5 mM
                            KCl:            50 mM
                            Tris-HCl:       10 mM
                            pH:             8.3.
FEATURES            Location/Qualifiers
     source         1. .425
                    /organism="Danio rerio"
                    /mol_type="genomic DNA"
                    /strain="AB"
                    /db_xref="taxon:7955"
                    /clone="Z14484"
                    /sex="F"
                    /clone_lib="Zebrafish AB"
                    /dev_stage="Adult"
                    /lab_host="DH5alphaF'IQ"
                    /note="Vector: m13MP19 with added BstXI site; V-type:
                    Phage; Genomic DNA  from a single adult Zebrafish of AB
                    strain was digested with  AluI,  Cac8I, HaeIII, NlaVI, or
                    RsaI.  Fragments in the range of 250-500 bp were  gel
                    purified and a BstX1 linker was added.  The fragments were
                    cloned into a  modified M13mp19 vector and transformed
                    into E. Coli DH5alpha.  Microsatelllite sequences were
                    screened with labeled d(CA)15 and d(GT)15 oligonucleotide
                    probes."
     STS            248. .399
     primer_bind    248. .267
     primer_bind    complement(380. .399)
ORIGIN

Alignment Scores:
Pred. No.:              103             Length:         425
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     10              Gaps:           0

US-09-858-332G-28 (1-6) x G46669 (1-425)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        169 CACAATCATAACCATAAC 186

RESULT 47
```

```
PFA430088
LOCUS       PFA430088               427 bp    DNA     linear   INV 07-FEB-2002
DEFINITION  Plasmodium falciparum partial rh1 gene, isolate FCB1.
ACCESSION   AJ430088
VERSION     AJ430088.1  GI:18642431
KEYWORDS    rh1 gene.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1
  AUTHORS   Taylor,H.M., Grainger,M. and Holder,A.A.
  TITLE     Variation in the expression of a Plasmodium falciparum protein
            family implicated in erythrocyte invasion
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 427)
  AUTHORS   Taylor,H.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-FEB-2002) Taylor H.M., Division of Parasitology,
            National Institute for Medical Research, Mill Hill, London, NW7
            1AA, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1. .427
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="FCB1"
                     /db_xref="taxon:5833"
                     /country="Colombia"
     gene            1. .426
                     /gene="rh1"
     CDS             <1. .>426
                     /gene="rh1"
                     /function="putative role in erythrocyte invasion"
                     /codon_start=1
                     /product="RH1 protein"
                     /protein_id="CAD23028.1"
                     /db_xref="GI:18642432"
                     /db_xref="UniProt/TrEMBL:Q8T8F1"
                     /translation="LENIKNNENIFDNIQLKKKDIDDIIININNTKETYLKELNKKKM
                     LQNKKKVDEKSEINNHHTLQHDNQNVEQKNKIKDHNLITKPNNNSSEESHQNEQMKEQ
                     NKNILEKQTRNIKPHHVHNHNHNHNHNHNHNHNHNQNQKD"
ORIGIN

Alignment Scores:
Pred. No.:              103             Length:         427
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-858-332G-28 (1-6) x PFA430088 (1-427)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        358 CATAATCATAATCATAAT 375

RESULT 48
BV211484
LOCUS       BV211484                431 bp    DNA     linear   STS 04-JAN-2005
DEFINITION  BARC1076 Wheat var Chinese Spring Genomic DNA Triticum aestivum STS
            genomic, sequence tagged site.
ACCESSION   BV211484
VERSION     BV211484.1  GI:56847231
KEYWORDS    STS.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 431)
  AUTHORS   Song,Q.J., Shi,J.R., Singh,S., Fickus,E.W., Costa,J.M., Lewis,J.,
            Gill,B.S., Ward,R. and Cregan,P.B.
  TITLE     Development and Mapping of Microsatellite (SSR) Markers in Wheat
```

```
  JOURNAL   Theor. Appl. Genet. (2005) In press
COMMENT     Synonyms: XBARC1076
            Contact: Dr. Perry Cregan
            Soybean Genomics Improvement Laboratory
            United States Department of Agriculture
            BARC-WEST, Bldg. 006, Rm 100, 10300 Baltimore Ave., MD 20705, USA
            Tel: 301-504-5070
            Fax: 301-504-5728
            Email: creganp@ba.ars.usda.gov
            Primer A: CGCTGGTCACACTATGGGTCAAT
            Primer B: CGCATAAACCAACTTCTCTCATC
            STS size: 156
            PCR Profile:
                    Presoak:          95 degrees C for 3.0 minute(s)
                    Denaturation:     95 degrees C for 45 second(s)
                    Annealing:        48 degrees C for 45 second(s)
                    Polymerization:   72 degrees C for 45 second(s)
                    PCR Cycles:       35
                    Thermal Cycler: Perkin Elmer TC or MJ PTC-225 or MBS
            Satellite 384
            Protocol:
                    Template:         30-100 ng
                    Primer:           each 1.5 uM
                    dNTPs:            each 200 uM
                    Taq Polymerase:   0.05 units/ul
                    Total Vol:        10 ul

            Buffer:
                    MgCl2:            2.5 mM
                    KCl:             50 mM
                    Tris-HCl:        10 mM
                    pH:               8.3.
FEATURES              Location/Qualifiers
     source           1. .431
                      /organism="Triticum aestivum"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:4565"
                      /clone_lib="Wheat var Chinese Spring Genomic DNA"
                      /note="Vector: pBluescript; V-type: Plasmid; Genomic
                      fragments cloned into pBluescript"
     STS              <1. .>431
ORIGIN

Alignment Scores:
Pred. No.:                104             Length:         431
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       10              Gaps:           0

US-09-858-332G-28 (1-6) x BV211484 (1-431)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        210 CATAATCATAATCATAAT 227

RESULT 49
AR547090/c
LOCUS       AR547090                 432 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 2221 from patent US 6747137.
ACCESSION   AR547090
VERSION     AR547090.1  GI:53940265
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 432)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 2221 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES              Location/Qualifiers
     source           1. .432
                      /organism="unknown"
                      /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:                104             Length:         432
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x AR547090 (1-432)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        386 CACAACCACAACCACAAC 369

RESULT 50
AY503292/c
LOCUS       AY503292                 441 bp    DNA     linear   PLN 16-MAR-2004
DEFINITION  Sorghum bicolor clone 0193-102069 RFLP probe pSB0193.
ACCESSION   AY503292
VERSION     AY503292.1  GI:45362150
KEYWORDS    .
SOURCE      Sorghum bicolor (sorghum)
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 441)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Comparative population genetics of the panicoid grasses: Sequence
            polymorphism, linkage disequilibrium, and selection in a diverse
            sample of Sorghum bicolor
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 441)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2003) Institute for Genomic Diversity, Cornell
            University, Biotechnology Building, Ithaca, NY 14853, USA
FEATURES              Location/Qualifiers
     source           1. .441
                      /organism="Sorghum bicolor"
                      /mol_type="genomic DNA"
                      /cultivar="race durra"
                      /db_xref="taxon:4558"
                      /clone="0193-102069"
     misc_feature     1. .441
                      /note="RFLP probe pSB0193"
ORIGIN

Alignment Scores:
Pred. No.:                106             Length:         441
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       15              Gaps:           0

US-09-858-332G-28 (1-6) x AY503292 (1-441)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        397 CACAACCACAACCACAAC 380

RESULT 51
G47457/c
LOCUS       G47457                   447 bp    DNA     linear   STS 23-MAR-1999
DEFINITION  Z24495_1 Zebrafish AB Danio rerio STS genomic clone Z24495 5',
            sequence tagged site.
```

```
ACCESSION   G47457
VERSION     G47457.1  GI:4493954
KEYWORDS    STS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 447)
  AUTHORS   Ziniti,J., Jackson,D.G. and Fishman,M.C.
  TITLE     A Zebrafish microsatellite map
  JOURNAL   Unpublished (1999)
COMMENT
            Contact: Mark C. Fishman
            Cardiovascular Research Center
            Massachusetts General Hospital
            Mail code 1494100A, 149 13th Street, Charlestown, MA 02129, USA
            Fax: 6177265806
            Email: fishman@mgh.cvrc.harvard.edu
            http://zebrafish.mgh.harvard.edu
            Primer A: GAACGTCCTTCATCCACCAC
            Primer B: TTGCACAAAGGTGTATGCGT
            STS size: 227
            PCR Profile:
                Presoak:        94 degrees C for 5.0 minutes
                Denaturation:   94 degrees C for 1.0 minute
                Annealing:      58 degrees C for 1.0 minute
                Polymerization: 72 degrees C for 1.5 minute
                PCR Cycles:     27
                Thermal Cycler: MJ Research PTC-100
            Protocol:
                Template:       10 ng
                Primer:         each 375 nM
                dNTPs:          each 200 uM
                Taq Polymerase: 0.034 units/ul
                Total Vol:      10 ul

            Buffer:
                            MgCL2:          1.5 mM
                            KCl:            50 mM
                            Tris-HCl:       10 mM
                            pH:             8.3.
FEATURES            Location/Qualifiers
     source         1. .447
                    /organism="Danio rerio"
                    /mol_type="genomic DNA"
                    /strain="AB"
                    /db_xref="taxon:7955"
                    /clone="Z24495"
                    /sex="F"
                    /clone_lib="Zebrafish AB"
                    /dev_stage="Adult"
                    /lab_host="DH5alphaF'IQ"
                    /note="Vector: m13MP19 with added BstXI site; V-type:
                    Phage; Genomic DNA  from a single adult Zebrafish of AB
                    strain was digested with  AluI,  Cac8I, HaeIII, NlaVI, or
                    RsaI.  Fragments in the range of 250-500 bp were  gel
                    purified and a BstX1 linker was added.  The fragments were
                    cloned into a  modified M13mp19 vector and transformed
                    into E. Coli DH5alpha.  Microsatelllite sequences were
                    screened with labeled d(CA)15 and d(GT)15 oligonucleotide
                    probes."
     STS            90. .316
     primer_bind    90. .109
     primer_bind    complement(297. .316)
ORIGIN

Alignment Scores:
Pred. No.:              108             Length:         447
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     10              Gaps:           0

US-09-858-332G-28 (1-6) x G47457 (1-447)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        198 CATAATCATAATCATAAT 181

RESULT 52
AY503290/c
LOCUS       AY503290                450 bp    DNA     linear   PLN 16-MAR-2004
DEFINITION  Sorghum bicolor clone 0193-BTx3197 RFLP probe pSB0193.
ACCESSION   AY503290
VERSION     AY503290.1  GI:45362148
KEYWORDS    .
SOURCE      Sorghum bicolor (sorghum)
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Comparative population genetics of the panicoid grasses: Sequence
            polymorphism, linkage disequilibrium, and selection in a diverse
            sample of Sorghum bicolor
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2003) Institute for Genomic Diversity, Cornell
            University, Biotechnology Building, Ithaca, NY 14853, USA
FEATURES            Location/Qualifiers
     source         1. .450
                    /organism="Sorghum bicolor"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:4558"
                    /clone="0193-BTx3197"
     misc_feature   1. .450
                    /note="RFLP probe pSB0193"
ORIGIN

Alignment Scores:
Pred. No.:              108             Length:         450
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x AY503290 (1-450)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        397 CACAACCACAACCACAAC 380

RESULT 53
AY503291/c
LOCUS       AY503291                450 bp    DNA     linear   PLN 16-MAR-2004
DEFINITION  Sorghum bicolor clone 0193-RTx430 RFLP probe pSB0193.
ACCESSION   AY503291
VERSION     AY503291.1  GI:45362149
KEYWORDS    .
SOURCE      Sorghum bicolor (sorghum)
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Comparative population genetics of the panicoid grasses: Sequence
            polymorphism, linkage disequilibrium, and selection in a diverse
            sample of Sorghum bicolor
  JOURNAL   Unpublished
```

```
REFERENCE   2  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2003) Institute for Genomic Diversity, Cornell
            University, Biotechnology Building, Ithaca, NY 14853, USA
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="Sorghum bicolor"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4558"
                     /clone="0193-RTx430"
     misc_feature    1. .450
                     /note="RFLP probe pSB0193"
ORIGIN

Alignment Scores:
Pred. No.:              108           Length:        450
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     15            Gaps:          0

US-09-858-332G-28 (1-6) x AY503291 (1-450)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        397 CACAACCACAACCACAAC 380

RESULT 54
AY503293/c
LOCUS       AY503293                450 bp    DNA     linear   PLN 16-MAR-2004
DEFINITION  Sorghum bicolor clone 0193-152702 RFLP probe pSB0193.
ACCESSION   AY503293
VERSION     AY503293.1  GI:45362151
KEYWORDS    .
SOURCE      Sorghum bicolor (sorghum)
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Comparative population genetics of the panicoid grasses: Sequence
            polymorphism, linkage disequilibrium, and selection in a diverse
            sample of Sorghum bicolor
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2003) Institute for Genomic Diversity, Cornell
            University, Biotechnology Building, Ithaca, NY 14853, USA
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="Sorghum bicolor"
                     /mol_type="genomic DNA"
                     /cultivar="race caudatum"
                     /db_xref="taxon:4558"
                     /clone="0193-152702"
     misc_feature    1. .450
                     /note="RFLP probe pSB0193"
ORIGIN

Alignment Scores:
Pred. No.:              108           Length:        450
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     15            Gaps:          0

US-09-858-332G-28 (1-6) x AY503293 (1-450)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        397 CACAACCACAACCACAAC 380

RESULT 55
AY503294/c
LOCUS       AY503294                450 bp    DNA     linear   PLN 16-MAR-2004
DEFINITION  Sorghum bicolor clone 0193-195684a RFLP probe pSB0193.
ACCESSION   AY503294
VERSION     AY503294.1  GI:45362152
KEYWORDS    .
SOURCE      Sorghum bicolor (sorghum)
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Comparative population genetics of the panicoid grasses: Sequence
            polymorphism, linkage disequilibrium, and selection in a diverse
            sample of Sorghum bicolor
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2003) Institute for Genomic Diversity, Cornell
            University, Biotechnology Building, Ithaca, NY 14853, USA
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="Sorghum bicolor"
                     /mol_type="genomic DNA"
                     /cultivar="race durra"
                     /db_xref="taxon:4558"
                     /clone="0193-195684a"
     misc_feature    1. .450
                     /note="RFLP probe pSB0193"
ORIGIN

Alignment Scores:
Pred. No.:              108           Length:        450
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     15            Gaps:          0

US-09-858-332G-28 (1-6) x AY503294 (1-450)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        397 CACAACCACAACCACAAC 380

RESULT 56
AY503295/c
LOCUS       AY503295                450 bp    DNA     linear   PLN 16-MAR-2004
DEFINITION  Sorghum bicolor clone 0193-195684b RFLP probe pSB0193.
ACCESSION   AY503295
VERSION     AY503295.1  GI:45362153
KEYWORDS    .
SOURCE      Sorghum bicolor (sorghum)
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Comparative population genetics of the panicoid grasses: Sequence
            polymorphism, linkage disequilibrium, and selection in a diverse
            sample of Sorghum bicolor
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (04-DEC-2003) Institute for Genomic Diversity, Cornell
            University, Biotechnology Building, Ithaca, NY 14853, USA
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="Sorghum bicolor"
                     /mol_type="genomic DNA"
                     /cultivar="race durra"
                     /db_xref="taxon:4558"
                     /clone="0193-195684b"
     misc_feature    1. .450
                     /note="RFLP probe pSB0193"
ORIGIN

Alignment Scores:
Pred. No.:              108           Length:         450
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     15            Gaps:           0

US-09-858-332G-28 (1-6) x AY503295 (1-450)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        397 CACAACCACAACCACAAC 380

RESULT 57
AY503296/c
LOCUS       AY503296                450 bp    DNA     linear   PLN 16-MAR-2004
DEFINITION  Sorghum arundinaceum clone 0193-225905 RFLP probe pSB0193.
ACCESSION   AY503296
VERSION     AY503296.1  GI:45362154
KEYWORDS    .
SOURCE      Sorghum arundinaceum
  ORGANISM  Sorghum arundinaceum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Comparative population genetics of the panicoid grasses: Sequence
            polymorphism, linkage disequilibrium, and selection in a diverse
            sample of Sorghum bicolor
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2003) Institute for Genomic Diversity, Cornell
            University, Biotechnology Building, Ithaca, NY 14853, USA
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="Sorghum arundinaceum"
                     /mol_type="genomic DNA"
                     /cultivar="race arundinaceum"
                     /db_xref="taxon:91525"
                     /clone="0193-225905"
     misc_feature    1. .450
                     /note="RFLP probe pSB0193"
ORIGIN

Alignment Scores:
Pred. No.:              108           Length:         450
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     15            Gaps:           0

US-09-858-332G-28 (1-6) x AY503296 (1-450)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        397 CACAACCACAACCACAAC 380

RESULT 58
AY503297/c
LOCUS       AY503297                450 bp    DNA     linear   PLN 16-MAR-2004
DEFINITION  Sorghum bicolor clone 0193-22913 RFLP probe pSB0193.
ACCESSION   AY503297
VERSION     AY503297.1  GI:45362155
KEYWORDS    .
SOURCE      Sorghum bicolor (sorghum)
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Comparative population genetics of the panicoid grasses: Sequence
            polymorphism, linkage disequilibrium, and selection in a diverse
            sample of Sorghum bicolor
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2003) Institute for Genomic Diversity, Cornell
            University, Biotechnology Building, Ithaca, NY 14853, USA
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="Sorghum bicolor"
                     /mol_type="genomic DNA"
                     /cultivar="race bicolor"
                     /db_xref="taxon:4558"
                     /clone="0193-22913"
     misc_feature    1. .450
                     /note="RFLP probe pSB0193"
ORIGIN

Alignment Scores:
Pred. No.:              108           Length:         450
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     15            Gaps:           0

US-09-858-332G-28 (1-6) x AY503297 (1-450)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        397 CACAACCACAACCACAAC 380

RESULT 59
AY503298/c
LOCUS       AY503298                450 bp    DNA     linear   PLN 16-MAR-2004
DEFINITION  Sorghum bicolor clone 0193-246712 RFLP probe pSB0193.
ACCESSION   AY503298
VERSION     AY503298.1  GI:45362156
KEYWORDS    .
SOURCE      Sorghum bicolor (sorghum)
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Comparative population genetics of the panicoid grasses: Sequence
            polymorphism, linkage disequilibrium, and selection in a diverse
            sample of Sorghum bicolor
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2003) Institute for Genomic Diversity, Cornell
            University, Biotechnology Building, Ithaca, NY 14853, USA
```

```
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="Sorghum bicolor"
                     /mol_type="genomic DNA"
                     /cultivar="race durra"
                     /db_xref="taxon:4558"
                     /clone="0193-246712"
     misc_feature    1. .450
                     /note="RFLP probe pSB0193"
ORIGIN

Alignment Scores:
Pred. No.:              108           Length:        450
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     15            Gaps:          0

US-09-858-332G-28 (1-6) x AY503298 (1-450)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        397 CACAACCACAACCACAAC 380

RESULT 60
AY503299/c
LOCUS       AY503299                450 bp    DNA     linear   PLN 16-MAR-2004
DEFINITION  Sorghum bicolor clone 0193-267380 RFLP probe pSB0193.
ACCESSION   AY503299
VERSION     AY503299.1  GI:45362157
KEYWORDS    .
SOURCE      Sorghum bicolor (sorghum)
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Comparative population genetics of the panicoid grasses: Sequence
            polymorphism, linkage disequilibrium, and selection in a diverse
            sample of Sorghum bicolor
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2003) Institute for Genomic Diversity, Cornell
            University, Biotechnology Building, Ithaca, NY 14853, USA
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="Sorghum bicolor"
                     /mol_type="genomic DNA"
                     /cultivar="race kafir"
                     /db_xref="taxon:4558"
                     /clone="0193-267380"
     misc_feature    1. .450
                     /note="RFLP probe pSB0193"
ORIGIN

Alignment Scores:
Pred. No.:              108           Length:        450
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     15            Gaps:          0

US-09-858-332G-28 (1-6) x AY503299 (1-450)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        397 CACAACCACAACCACAAC 380

RESULT 61
AY503302/c
LOCUS       AY503302                450 bp    DNA     linear   PLN 16-MAR-2004
DEFINITION  Sorghum bicolor clone 0193-514605 RFLP probe pSB0193.
ACCESSION   AY503302
VERSION     AY503302.1  GI:45362160
KEYWORDS    .
SOURCE      Sorghum bicolor (sorghum)
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Comparative population genetics of the panicoid grasses: Sequence
            polymorphism, linkage disequilibrium, and selection in a diverse
            sample of Sorghum bicolor
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2003) Institute for Genomic Diversity, Cornell
            University, Biotechnology Building, Ithaca, NY 14853, USA
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="Sorghum bicolor"
                     /mol_type="genomic DNA"
                     /cultivar="race caudatum"
                     /db_xref="taxon:4558"
                     /clone="0193-514605"
     misc_feature    1. .450
                     /note="RFLP probe pSB0193"
ORIGIN

Alignment Scores:
Pred. No.:              108           Length:        450
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     15            Gaps:          0

US-09-858-332G-28 (1-6) x AY503302 (1-450)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        397 CACAACCACAACCACAAC 380

RESULT 62
AY503304/c
LOCUS       AY503304                450 bp    DNA     linear   PLN 16-MAR-2004
DEFINITION  Sorghum bicolor clone 0193-56174 RFLP probe pSB0193.
ACCESSION   AY503304
VERSION     AY503304.1  GI:45362162
KEYWORDS    .
SOURCE      Sorghum bicolor (sorghum)
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Comparative population genetics of the panicoid grasses: Sequence
            polymorphism, linkage disequilibrium, and selection in a diverse
            sample of Sorghum bicolor
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2003) Institute for Genomic Diversity, Cornell
            University, Biotechnology Building, Ithaca, NY 14853, USA
FEATURES             Location/Qualifiers
     source          1. .450
```

```
                     /organism="Sorghum bicolor"
                     /mol_type="genomic DNA"
                     /cultivar="race kafir"
                     /db_xref="taxon:4558"
                     /clone="0193-56174"
     misc_feature    1. .450
                     /note="RFLP probe pSB0193"
ORIGIN

Alignment Scores:
Pred. No.:                108             Length:         450
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       15              Gaps:           0

US-09-858-332G-28 (1-6) x AY503304 (1-450)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        397 CACAACCACAACCACAAC 380

RESULT 63
AY503305/c
LOCUS       AY503305                 450 bp    DNA     linear   PLN 16-MAR-2004
DEFINITION  Sorghum bicolor clone 0193-77034 RFLP probe pSB0193.
ACCESSION   AY503305
VERSION     AY503305.1  GI:45362163
KEYWORDS    .
SOURCE      Sorghum bicolor (sorghum)
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Comparative population genetics of the panicoid grasses: Sequence
            polymorphism, linkage disequilibrium, and selection in a diverse
            sample of Sorghum bicolor
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2003) Institute for Genomic Diversity, Cornell
            University, Biotechnology Building, Ithaca, NY 14853, USA
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="Sorghum bicolor"
                     /mol_type="genomic DNA"
                     /cultivar="race kafir"
                     /db_xref="taxon:4558"
                     /clone="0193-77034"
     misc_feature    1. .450
                     /note="RFLP probe pSB0193"
ORIGIN

Alignment Scores:
Pred. No.:                108             Length:         450
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       15              Gaps:           0

US-09-858-332G-28 (1-6) x AY503305 (1-450)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        397 CACAACCACAACCACAAC 380

RESULT 64
AY503306/c
LOCUS       AY503306                 450 bp    DNA     linear   PLN 16-MAR-2004
DEFINITION  Sorghum bicolor clone 0193-82459 RFLP probe pSB0193.
ACCESSION   AY503306
VERSION     AY503306.1  GI:45362164
KEYWORDS    .
SOURCE      Sorghum bicolor (sorghum)
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Comparative population genetics of the panicoid grasses: Sequence
            polymorphism, linkage disequilibrium, and selection in a diverse
            sample of Sorghum bicolor
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2003) Institute for Genomic Diversity, Cornell
            University, Biotechnology Building, Ithaca, NY 14853, USA
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="Sorghum bicolor"
                     /mol_type="genomic DNA"
                     /cultivar="race bicolor"
                     /db_xref="taxon:4558"
                     /clone="0193-82459"
     misc_feature    1. .450
                     /note="RFLP probe pSB0193"
ORIGIN

Alignment Scores:
Pred. No.:                108             Length:         450
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       15              Gaps:           0

US-09-858-332G-28 (1-6) x AY503306 (1-450)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        397 CACAACCACAACCACAAC 380

RESULT 65
AY503308/c
LOCUS       AY503308                 450 bp    DNA     linear   PLN 16-MAR-2004
DEFINITION  Sorghum arundinaceum clone 0193-LWA4 RFLP probe pSB0193.
ACCESSION   AY503308
VERSION     AY503308.1  GI:45362166
KEYWORDS    .
SOURCE      Sorghum arundinaceum
  ORGANISM  Sorghum arundinaceum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Comparative population genetics of the panicoid grasses: Sequence
            polymorphism, linkage disequilibrium, and selection in a diverse
            sample of Sorghum bicolor
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2003) Institute for Genomic Diversity, Cornell
            University, Biotechnology Building, Ithaca, NY 14853, USA
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="Sorghum arundinaceum"
                     /mol_type="genomic DNA"
```

```
                     /cultivar="race verticilliflorum"
                     /db_xref="taxon:91525"
                     /clone="0193-LWA4"
     misc_feature    1. .450
                     /note="RFLP probe pSB0193"
ORIGIN

Alignment Scores:
Pred. No.:              108           Length:        450
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     15            Gaps:          0

US-09-858-332G-28 (1-6) x AY503308 (1-450)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        397 CACAACCACAACCACAAC 380

RESULT 66
AY503309/c
LOCUS       AY503309                450 bp    DNA     linear   PLN 16-MAR-2004
DEFINITION  Sorghum arundinaceum clone 0193-LWA8 RFLP probe pSB0193.
ACCESSION   AY503309
VERSION     AY503309.1  GI:45362167
KEYWORDS    .
SOURCE      Sorghum arundinaceum
  ORGANISM  Sorghum arundinaceum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Comparative population genetics of the panicoid grasses: Sequence
            polymorphism, linkage disequilibrium, and selection in a diverse
            sample of Sorghum bicolor
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2003) Institute for Genomic Diversity, Cornell
            University, Biotechnology Building, Ithaca, NY 14853, USA
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="Sorghum arundinaceum"
                     /mol_type="genomic DNA"
                     /cultivar="race aethiopicum"
                     /db_xref="taxon:91525"
                     /clone="0193-LWA8"
     misc_feature    1. .450
                     /note="RFLP probe pSB0193"
ORIGIN

Alignment Scores:
Pred. No.:              108           Length:        450
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     15            Gaps:          0

US-09-858-332G-28 (1-6) x AY503309 (1-450)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        397 CACAACCACAACCACAAC 380

RESULT 67
AY503310/c
LOCUS       AY503310                450 bp    DNA     linear   PLN 16-MAR-2004
DEFINITION  Sorghum bicolor clone 0193-RTx406 RFLP probe pSB0193.
ACCESSION   AY503310
VERSION     AY503310.1  GI:45362168
KEYWORDS    .
SOURCE      Sorghum bicolor (sorghum)
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Comparative population genetics of the panicoid grasses: Sequence
            polymorphism, linkage disequilibrium, and selection in a diverse
            sample of Sorghum bicolor
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2003) Institute for Genomic Diversity, Cornell
            University, Biotechnology Building, Ithaca, NY 14853, USA
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="Sorghum bicolor"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4558"
                     /clone="0193-RTx406"
     misc_feature    1. .450
                     /note="RFLP probe pSB0193"
ORIGIN

Alignment Scores:
Pred. No.:              108           Length:        450
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     15            Gaps:          0

US-09-858-332G-28 (1-6) x AY503310 (1-450)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        397 CACAACCACAACCACAAC 380

RESULT 68
AY503311/c
LOCUS       AY503311                450 bp    DNA     linear   PLN 16-MAR-2004
DEFINITION  Sorghum bicolor clone 0193-SC0033 RFLP probe pSB0193.
ACCESSION   AY503311
VERSION     AY503311.1  GI:45362169
KEYWORDS    .
SOURCE      Sorghum bicolor (sorghum)
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Comparative population genetics of the panicoid grasses: Sequence
            polymorphism, linkage disequilibrium, and selection in a diverse
            sample of Sorghum bicolor
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2003) Institute for Genomic Diversity, Cornell
            University, Biotechnology Building, Ithaca, NY 14853, USA
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="Sorghum bicolor"
                     /mol_type="genomic DNA"
                     /cultivar="race durra"
                     /db_xref="taxon:4558"
                     /clone="0193-SC0033"
```

```
     misc_feature    1. .450
                     /note="RFLP probe pSB0193"
ORIGIN

Alignment Scores:
Pred. No.:                108             Length:         450
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       15              Gaps:           0

US-09-858-332G-28 (1-6) x AY503311 (1-450)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        397 CACAACCACAACCACAAC 380

RESULT 69
AY503312/c
LOCUS       AY503312                450 bp    DNA     linear   PLN 16-MAR-2004
DEFINITION  Sorghum bicolor clone 0193-SC0155 RFLP probe pSB0193.
ACCESSION   AY503312
VERSION     AY503312.1  GI:45362170
KEYWORDS    .
SOURCE      Sorghum bicolor (sorghum)
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Comparative population genetics of the panicoid grasses: Sequence
            polymorphism, linkage disequilibrium, and selection in a diverse
            sample of Sorghum bicolor
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2003) Institute for Genomic Diversity, Cornell
            University, Biotechnology Building, Ithaca, NY 14853, USA
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="Sorghum bicolor"
                     /mol_type="genomic DNA"
                     /cultivar="race durra-bicolor"
                     /db_xref="taxon:4558"
                     /clone="0193-SC0155"
     misc_feature    1. .450
                     /note="RFLP probe pSB0193"
ORIGIN

Alignment Scores:
Pred. No.:                108             Length:         450
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       15              Gaps:           0

US-09-858-332G-28 (1-6) x AY503312 (1-450)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        397 CACAACCACAACCACAAC 380

RESULT 70
AY503313/c
LOCUS       AY503313                450 bp    DNA     linear   PLN 16-MAR-2004
DEFINITION  Sorghum bicolor clone 0193-SC0326 RFLP probe pSB0193.
ACCESSION   AY503313
VERSION     AY503313.1  GI:45362171
KEYWORDS    .
SOURCE      Sorghum bicolor (sorghum)
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Comparative population genetics of the panicoid grasses: Sequence
            polymorphism, linkage disequilibrium, and selection in a diverse
            sample of Sorghum bicolor
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2003) Institute for Genomic Diversity, Cornell
            University, Biotechnology Building, Ithaca, NY 14853, USA
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="Sorghum bicolor"
                     /mol_type="genomic DNA"
                     /cultivar="race caudatum"
                     /db_xref="taxon:4558"
                     /clone="0193-SC0326"
     misc_feature    1. .450
                     /note="RFLP probe pSB0193"
ORIGIN

Alignment Scores:
Pred. No.:                108             Length:         450
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       15              Gaps:           0

US-09-858-332G-28 (1-6) x AY503313 (1-450)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        397 CACAACCACAACCACAAC 380

RESULT 71
AY503314/c
LOCUS       AY503314                450 bp    DNA     linear   PLN 16-MAR-2004
DEFINITION  Sorghum bicolor clone 0193-SC0599 RFLP probe pSB0193.
ACCESSION   AY503314
VERSION     AY503314.1  GI:45362172
KEYWORDS    .
SOURCE      Sorghum bicolor (sorghum)
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Comparative population genetics of the panicoid grasses: Sequence
            polymorphism, linkage disequilibrium, and selection in a diverse
            sample of Sorghum bicolor
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2003) Institute for Genomic Diversity, Cornell
            University, Biotechnology Building, Ithaca, NY 14853, USA
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="Sorghum bicolor"
                     /mol_type="genomic DNA"
                     /cultivar="race caudatum"
                     /db_xref="taxon:4558"
                     /clone="0193-SC0599"
     misc_feature    1. .450
                     /note="RFLP probe pSB0193"
```

```
ORIGIN

Alignment Scores:
Pred. No.:              108           Length:        450
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     15            Gaps:          0

US-09-858-332G-28 (1-6) x AY503314 (1-450)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        397 CACAACCACAACCACAAC 380

RESULT 72
AY503315/c
LOCUS       AY503315                450 bp    DNA     linear   PLN 16-MAR-2004
DEFINITION  Sorghum bicolor clone 0193-SC0706 RFLP probe pSB0193.
ACCESSION   AY503315
VERSION     AY503315.1  GI:45362173
KEYWORDS    .
SOURCE      Sorghum bicolor (sorghum)
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Comparative population genetics of the panicoid grasses: Sequence
            polymorphism, linkage disequilibrium, and selection in a diverse
            sample of Sorghum bicolor
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 450)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2003) Institute for Genomic Diversity, Cornell
            University, Biotechnology Building, Ithaca, NY 14853, USA
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="Sorghum bicolor"
                     /mol_type="genomic DNA"
                     /cultivar="race caudatum"
                     /db_xref="taxon:4558"
                     /clone="0193-SC0706"
     misc_feature    1. .450
                     /note="RFLP probe pSB0193"
ORIGIN

Alignment Scores:
Pred. No.:              108           Length:        450
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     15            Gaps:          0

US-09-858-332G-28 (1-6) x AY503315 (1-450)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        397 CACAACCACAACCACAAC 380

RESULT 73
AY503300/c
LOCUS       AY503300                451 bp    DNA     linear   PLN 16-MAR-2004
DEFINITION  Sorghum bicolor clone 0193-50875 RFLP probe pSB0193.
ACCESSION   AY503300
VERSION     AY503300.1  GI:45362158
KEYWORDS    .
SOURCE      Sorghum bicolor (sorghum)
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 451)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Comparative population genetics of the panicoid grasses: Sequence
            polymorphism, linkage disequilibrium, and selection in a diverse
            sample of Sorghum bicolor
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 451)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2003) Institute for Genomic Diversity, Cornell
            University, Biotechnology Building, Ithaca, NY 14853, USA
FEATURES             Location/Qualifiers
     source          1. .451
                     /organism="Sorghum bicolor"
                     /mol_type="genomic DNA"
                     /cultivar="race guinea"
                     /db_xref="taxon:4558"
                     /clone="0193-50875"
     misc_feature    1. .451
                     /note="RFLP probe pSB0193"
ORIGIN

Alignment Scores:
Pred. No.:              109           Length:        451
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     15            Gaps:          0

US-09-858-332G-28 (1-6) x AY503300 (1-451)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        397 CACAACCACAACCACAAC 380

RESULT 74
AY503303/c
LOCUS       AY503303                451 bp    DNA     linear   PLN 16-MAR-2004
DEFINITION  Sorghum bicolor clone 0193-56003 RFLP probe pSB0193.
ACCESSION   AY503303
VERSION     AY503303.1  GI:45362161
KEYWORDS    .
SOURCE      Sorghum bicolor (sorghum)
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 451)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Comparative population genetics of the panicoid grasses: Sequence
            polymorphism, linkage disequilibrium, and selection in a diverse
            sample of Sorghum bicolor
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 451)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2003) Institute for Genomic Diversity, Cornell
            University, Biotechnology Building, Ithaca, NY 14853, USA
FEATURES             Location/Qualifiers
     source          1. .451
                     /organism="Sorghum bicolor"
                     /mol_type="genomic DNA"
                     /cultivar="race bicolor"
                     /db_xref="taxon:4558"
                     /clone="0193-56003"
     misc_feature    1. .451
                     /note="RFLP probe pSB0193"
ORIGIN
```

```
Alignment Scores:
Pred. No.:                109             Length:         451
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       15              Gaps:           0

US-09-858-332G-28 (1-6) x AY503303 (1-451)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        397 CACAACCACAACCACAAC 380

RESULT 75
AY503307/c
LOCUS       AY503307                 451 bp    DNA     linear   PLN 16-MAR-2004
DEFINITION  Sorghum bicolor clone 0193-92381 RFLP probe pSB0193.
ACCESSION   AY503307
VERSION     AY503307.1  GI:45362165
KEYWORDS    .
SOURCE      Sorghum bicolor (sorghum)
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 451)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Comparative population genetics of the panicoid grasses: Sequence
            polymorphism, linkage disequilibrium, and selection in a diverse
            sample of Sorghum bicolor
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 451)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2003) Institute for Genomic Diversity, Cornell
            University, Biotechnology Building, Ithaca, NY 14853, USA
FEATURES             Location/Qualifiers
     source          1. .451
                     /organism="Sorghum bicolor"
                     /mol_type="genomic DNA"
                     /cultivar="race guinea"
                     /db_xref="taxon:4558"
                     /clone="0193-92381"
     misc_feature    1. .451
                     /note="RFLP probe pSB0193"
ORIGIN

Alignment Scores:
Pred. No.:                109             Length:         451
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       15              Gaps:           0

US-09-858-332G-28 (1-6) x AY503307 (1-451)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        397 CACAACCACAACCACAAC 380

RESULT 76
AR508437
LOCUS       AR508437                 456 bp    DNA     linear   PAT 22-SEP-2004
DEFINITION  Sequence 13397 from patent US 6703491.
ACCESSION   AR508437
VERSION     AR508437.1  GI:52443912
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 456)
  AUTHORS   Homburger,S.A., Ebens,A.J. Jr., Erickson,C.S., Francis-Lang,H.L.,
            Margolis,J.S., Reddy,B.P., Ruddy,D.A. and Buchman,A.R.
  TITLE     Drosophila sequences
  JOURNAL   Patent: US 6703491-A 13397 09-MAR-2004;
            Exelixis, Inc.; South San Francisco, CA
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:                110             Length:         456
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x AR508437 (1-456)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        291 CATAATCACAACCATAAT 308

RESULT 77
AY503301/c
LOCUS       AY503301                 456 bp    DNA     linear   PLN 16-MAR-2004
DEFINITION  Sorghum bicolor clone 0193-51030 RFLP probe pSB0193.
ACCESSION   AY503301
VERSION     AY503301.1  GI:45362159
KEYWORDS    .
SOURCE      Sorghum bicolor (sorghum)
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 456)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Comparative population genetics of the panicoid grasses: Sequence
            polymorphism, linkage disequilibrium, and selection in a diverse
            sample of Sorghum bicolor
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 456)
  AUTHORS   Hamblin,M.T., Mitchell,S.E. and Kresovich,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2003) Institute for Genomic Diversity, Cornell
            University, Biotechnology Building, Ithaca, NY 14853, USA
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="Sorghum bicolor"
                     /mol_type="genomic DNA"
                     /cultivar="race guinea"
                     /db_xref="taxon:4558"
                     /clone="0193-51030"
     misc_feature    1. .456
                     /note="RFLP probe pSB0193"
ORIGIN

Alignment Scores:
Pred. No.:                110             Length:         456
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       15              Gaps:           0

US-09-858-332G-28 (1-6) x AY503301 (1-456)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        403 CACAACCACAACCACAAC 386
```

```
RESULT 78
AR500762
LOCUS       AR500762                457 bp    DNA     linear   PAT 22-SEP-2004
DEFINITION  Sequence 5722 from patent US 6703491.
ACCESSION   AR500762
VERSION     AR500762.1  GI:52436237
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 457)
  AUTHORS   Homburger,S.A., Ebens,A.J. Jr., Erickson,C.S., Francis-Lang,H.L.,
            Margolis,J.S., Reddy,B.P., Ruddy,D.A. and Buchman,A.R.
  TITLE     Drosophila sequences
  JOURNAL   Patent: US 6703491-A 5722 09-MAR-2004;
            Exelixis, Inc.; South San Francisco, CA
FEATURES             Location/Qualifiers
     source          1. .457
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              110           Length:        457
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-28 (1-6) x AR500762 (1-457)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         15 CACAACCACAACCACAAC 32

RESULT 79
AR516044
LOCUS       AR516044                457 bp    DNA     linear   PAT 22-SEP-2004
DEFINITION  Sequence 21004 from patent US 6703491.
ACCESSION   AR516044
VERSION     AR516044.1  GI:52451519
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 457)
  AUTHORS   Homburger,S.A., Ebens,A.J. Jr., Erickson,C.S., Francis-Lang,H.L.,
            Margolis,J.S., Reddy,B.P., Ruddy,D.A. and Buchman,A.R.
  TITLE     Drosophila sequences
  JOURNAL   Patent: US 6703491-A 21004 09-MAR-2004;
            Exelixis, Inc.; South San Francisco, CA
FEATURES             Location/Qualifiers
     source          1. .457
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              110           Length:        457
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-28 (1-6) x AR516044 (1-457)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         15 CACAACCACAACCACAAC 32

RESULT 80
BV331153/c
LOCUS       BV331153                458 bp    DNA     linear   STS 26-JAN-2005
DEFINITION  S241P619RH11.T0 IndiaGrayWolf Canis lupus STS genomic, sequence
            tagged site.
ACCESSION   BV331153
VERSION     BV331153.1  GI:57531456
KEYWORDS    STS.
SOURCE      Canis lupus (gray wolf)
  ORGANISM  Canis lupus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
            Canis.
REFERENCE   1  (bases 1 to 458)
  AUTHORS   Lindblad-Toh,K.
  TITLE     The genome sequence of Canis familiaris
  JOURNAL   Unpublished (2004)
COMMENT
            Contact: Kerstin Lindblad-Toh
            Whitehead Institute for Biomedical Research, Center for Genome
            Research
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172521477
            Fax: 6172580903
            Email: kersli@genome.wi.mit.edu
            Primer A: No sequence submitted
            Primer B: No sequence submitted
            STS size: 458
            Protocol:
            WGS-discovery (WGS):
            Paired-end low-coverage whole genome shotgun reads were generated
            from 9 breeds
            (German Shepherd, Rottweiler, Bedlington Terrier, Beagle, Labrador
            Retriever, English
            Shepherd, Italian Greyhound, Alaskan Malamute and the Portuguese
            Water Dog -100,000 each)
            and five other canids (Chinese, Alaskan, Indian and Spanish Gray
            Wolf as well as the
            Californian Coyote).
            The WGS reads were placed uniquely on the CanFam1.0 boxer assembly
            and SNP detection was
            carried out by SSAHA-SNP. 863872 reads were annotated as STSs and
            485941 SNPs were
            annotated with alleles from the boxer and the breed or canid from
            which the particular
            read came. The validation rate for these SNPs was estimated at
            approximately 98%.
            WGA-discovery (WGA) of Boxer/Poodle SNPs:
            A second set of SNPs was generated using a similar methodology
            except that the contigs
            from the 1.5x poodle assembly (Kirkness 2003) were used instead of
            WGS reads. Since this
            sequence lacked base quality scores, arbitrary quality scores of
            phred 40 were assigned
            before the poddle seqeunce was placed uniquely on the CanFam1.0
            boxer assembly and SNP
            detection was carried out by SSAHA-SNP. 1637780 SNPs were annotated
            with alleles from the
            boxer and the poodle. The validation rate for these SNPs was
            estimated at approximately TBD%.
            Internal-WGA-discovery (I-WGA):
            A third set of SNPs were discovered by comparing reads in the WGA
            assembly. SNPs were
            defined as mismatch positions that had a base quality of >= 30 on
            both reads in a region
            that aligned without gaps, and with at most one additional mismatch
            in the ten flanking
            bases. For each allele, at least one additional read had to confirm
            it. 731476 SNPs were
            annotated with alleles between the two boxer alleles. The
            validation rate for these SNPs
            was estimated at approximately TBD%.
FEATURES             Location/Qualifiers
     source          1. .458
```

```
                     /organism="Canis lupus"
                     /mol_type="genomic DNA"
                     /strain="IndiaGrayWolf"
                     /db_xref="taxon:9612"
                     /map="+ 7 27-408 15324247-15323866"
                     /clone_lib="IndiaGrayWolf"
     STS             <1. .>458
ORIGIN

Alignment Scores:
Pred. No.:                110             Length:         458
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       10              Gaps:           0

US-09-858-332G-28 (1-6) x BV331153 (1-458)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         32 CATAATCATAATCATAAT 15

RESULT 81
DVNEURS1
LOCUS       DVNEURS1                  465 bp    DNA     linear   INV 23-AUG-2002
DEFINITION  Drosophila virilis neuralized protein (neu) gene, exon 1.
ACCESSION   U12591
VERSION     U12591.1  GI:535310
KEYWORDS    .
SEGMENT     1 of 3
SOURCE      Drosophila virilis
  ORGANISM  Drosophila virilis
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 465)
  AUTHORS   Zhou,L. and Boulianne,G.L.
  TITLE     Comparison of the neuralized genes of Drosophila virilis and D.
            melanogaster
  JOURNAL   Genome 37 (5), 840-847 (1994)
   PUBMED   8001814
REFERENCE   2  (bases 1 to 465)
  AUTHORS   Boulianne,G.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-1994) Gabrielle L. Boulianne, Physiology,
            University of Toronto, 6 Queens Park Crescent, Toronto, Ontario M5S
            1A8, Canada
FEATURES             Location/Qualifiers
     source          1. .465
                     /organism="Drosophila virilis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7244"
                     /clone="dv5"
                     /clone_lib="Drosophila virilis sau3a genomic library; S.J.
                     Neufeld et al., J. Mol. Evol. 32:415-420(1991)."
     exon            <1. .332
                     /gene="neu"
                     /number=1
ORIGIN

Alignment Scores:
Pred. No.:                112             Length:         465
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       2               Gaps:           0

US-09-858-332G-28 (1-6) x DVNEURS1 (1-465)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        173 CACAATCATAATCATAAT 190

RESULT 82
BV004681
LOCUS       BV004681                  467 bp    DNA     linear   STS 07-SEP-2002
DEFINITION  S208P6706FA10.T0 129S1/SvImJ Mus musculus STS genomic, sequence
            tagged site.
ACCESSION   BV004681
VERSION     BV004681.1  GI:22757757
KEYWORDS    STS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 467)
  AUTHORS   Wade,C.
  TITLE     Polymorphism Structure in the Mouse
  JOURNAL   Unpublished (2002)
COMMENT
            Contact: Kerstin Lindblad-Toh
            Whitehead Institute for Biomedical Research, Center for Genome
            Research
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172521477
            Fax: 6172580903
            Email: kersli@genome.wi.mit.edu
            Primer A: None
            Primer B: None
            STS size: 467
            Protocol:
            WGS-discovery: Paired-end low-coverage whole genome shotgun reads
            were generated from 129S1/SvImJ, C3H/HeJ, and BALB/cByJ. The WGS
            reads were placed uniquely on the MGSCv3 C57BL/6J assembly and SNP
            detection was carried out by SSAHA-SNP. 225,000 reads were
            annotated
            as STSs and 81,000 SNPs were annotated with alleles from C57BL/6J
            and the strain from which the particular read came. The validation
            rate for these SNPs was estimated at approximately 98%.
FEATURES             Location/Qualifiers
     source          1. .467
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="129S1/SvImJ"
                     /db_xref="taxon:10090"
                     /map="+ 9 11-411 114482865-114483253"
                     /clone_lib="129S1/SvImJ"
     STS             <1. .>467
ORIGIN

Alignment Scores:
Pred. No.:                112             Length:         467
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       10              Gaps:           0

US-09-858-332G-28 (1-6) x BV004681 (1-467)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        120 CACAACCACAACCACAAC 137

RESULT 83
BV213506/c
LOCUS       BV213506                  477 bp    DNA     linear   STS 07-JAN-2005
DEFINITION  S233P636FA3.T0 LabradorRetriever Canis familiaris STS genomic,
            sequence tagged site.
ACCESSION   BV213506
VERSION     BV213506.1  GI:57211321
KEYWORDS    STS.
SOURCE      Canis familiaris (dog)
```

```
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
            Canis.
REFERENCE   1  (bases 1 to 477)
  AUTHORS   Lindblad-Toh,K.
  TITLE     The genome sequence of Canis familiaris
  JOURNAL   Unpublished (2004)
COMMENT
            Contact: Kerstin Lindblad-Toh
            Whitehead Institute for Biomedical Research, Center for Genome
            Research
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172521477
            Fax: 6172580903
            Email: kersli@genome.wi.mit.edu
            Primer A: No sequence submitted
            Primer B: No sequence submitted
            STS size: 477
            Protocol:
            WGS-discovery (WGS):
            Paired-end low-coverage whole genome shotgun reads were generated
            from 9 breeds
            (German Shepherd, Rottweiler, Bedlington Terrier, Beagle, Labrador
            Retriever, English
            Shepherd, Italian Greyhound, Alaskan Malamute and the Portuguese
            Water Dog -100,000 each)
            and five other canids (Chinese, Alaskan, Indian and Spanish Gray
            Wolf as well as the
            Californian Coyote).
            The WGS reads were placed uniquely on the CanFam1.0 boxer assembly
            and SNP detection was
            carried out by SSAHA-SNP. 863872 reads were annotated as STSs and
            485941 SNPs were
            annotated with alleles from the boxer and the breed or canid from
            which the particular
            read came. The validation rate for these SNPs was estimated at
            approximately 98%.
            WGA-discovery (WGA) of Boxer/Poodle SNPs:
            A second set of SNPs was generated using a similar methodology
            except that the contigs
            from the 1.5x poodle assembly (Kirkness 2003) were used instead of
            WGS reads. Since this
            sequence lacked base quality scores, arbitrary quality scores of
            phred 40 were assigned
            before the poddle seqeunce was placed uniquely on the CanFam1.0
            boxer assembly and SNP
            detection was carried out by SSAHA-SNP. 1637780 SNPs were annotated
            with alleles from the
            boxer and the poodle. The validation rate for these SNPs was
            estimated at approximately TBD%.
            Internal-WGA-discovery (I-WGA):
            A third set of SNPs were discovered by comparing reads in the WGA
            assembly. SNPs were
            defined as mismatch positions that had a base quality of >= 30 on
            both reads in a region
            that aligned without gaps, and with at most one additional mismatch
            in the ten flanking
            bases. For each allele, at least one additional read had to confirm
            it. 731476 SNPs were
            annotated with alleles between the two boxer alleles. The
            validation rate for these SNPs
            was estimated at approximately TBD%.
FEATURES              Location/Qualifiers
     source           1. .477
                      /organism="Canis familiaris"
                      /mol_type="genomic DNA"
                      /strain="LabradorRetriever"
                      /db_xref="taxon:9615"
                      /map="+ 1 22-430 110323686-110323278"
                      /clone_lib="LabradorRetriever"
     STS              <1. .>477
ORIGIN
```

```
Alignment Scores:
Pred. No.:                115             Length:         477
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       10              Gaps:           0

US-09-858-332G-28 (1-6) x BV213506 (1-477)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        426 CACAACCACAACCACAAC 409

RESULT 84
BTU49779/c
LOCUS       BTU49779                 478 bp    mRNA    linear   MAM 16-AUG-2001
DEFINITION  Bos taurus immunoglobulin rearranged heavy chain variable region
            mRNA, partial cds.
ACCESSION   U49779
VERSION     U49779.1  GI:1293603
KEYWORDS    .
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 478)
  AUTHORS   Berens,S.J., Wylie,D.E. and Lopez,O.J.
  TITLE     Use of a single VH family and long CDR3s in the variable region of
            cattle Ig heavy chains
  JOURNAL   Int. Immunol. 9 (1), 189-199 (1997)
   PUBMED   9043960
REFERENCE   2  (bases 1 to 478)
  AUTHORS   Berens,S.J., Wylie,D. and Lopez,O.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-1996) Osvaldo J. Lopez, Biological Sciences,
            University of Nebraska, 325 Manter Hall, Lincoln, NE 68588-0118,
            USA
FEATURES              Location/Qualifiers
     source           1. .478
                      /organism="Bos taurus"
                      /mol_type="mRNA"
                      /db_xref="taxon:9913"
                      /cell_type="heterohybridoma"
                      /tissue_type="spleen"
                      /dev_stage="150 day fetus"
     CDS              <1. .>478
                      /codon_start=2
                      /product="immunoglobulin heavy chain variable region"
                      /protein_id="AAA98660.1"
                      /db_xref="GI:1293604"
                      /translation="LLGVVTLICSKMNPPWTLLFVLSAPRGVLSQVQLRESGPSLVKP
                      SQTLSLTCTVSGFSLSDNSVGWVRQPPGKALEWLGVIYAGGSTGYNPALKSRLSITKD
                      NSKSQVSLSLSSVTTEDTATYYCARVGFCYSGYGYGYGCGYGYGDAWGQGLLVTVSS"
ORIGIN

Alignment Scores:
Pred. No.:                115             Length:         478
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       4               Gaps:           0

US-09-858-332G-28 (1-6) x BTU49779 (1-478)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        437 CATAACCATAACCACAAC 420

RESULT 85
```

```
BV281117
LOCUS       BV281117                478 bp    DNA     linear   STS 22-JAN-2005
DEFINITION  S232P6238FE7.T0 Beagle Canis familiaris STS genomic, sequence
            tagged site.
ACCESSION   BV281117
VERSION     BV281117.1  GI:57443346
KEYWORDS    STS.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
            Canis.
REFERENCE   1  (bases 1 to 478)
  AUTHORS   Lindblad-Toh,K.
  TITLE     The genome sequence of Canis familiaris
  JOURNAL   Unpublished (2004)
COMMENT
            Contact: Kerstin Lindblad-Toh
            Whitehead Institute for Biomedical Research, Center for Genome
            Research
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172521477
            Fax: 6172580903
            Email: kersli@genome.wi.mit.edu
            Primer A: No sequence submitted
            Primer B: No sequence submitted
            STS size: 478
            Protocol:
            WGS-discovery (WGS):
            Paired-end low-coverage whole genome shotgun reads were generated
            from 9 breeds
            (German Shepherd, Rottweiler, Bedlington Terrier, Beagle, Labrador
            Retriever, English
            Shepherd, Italian Greyhound, Alaskan Malamute and the Portuguese
            Water Dog -100,000 each)
            and five other canids (Chinese, Alaskan, Indian and Spanish Gray
            Wolf as well as the
            Californian Coyote).
            The WGS reads were placed uniquely on the CanFam1.0 boxer assembly
            and SNP detection was
            carried out by SSAHA-SNP. 863872 reads were annotated as STSs and
            485941 SNPs were
            annotated with alleles from the boxer and the breed or canid from
            which the particular
            read came. The validation rate for these SNPs was estimated at
            approximately 98%.
            WGA-discovery (WGA) of Boxer/Poodle SNPs:
            A second set of SNPs was generated using a similar methodology
            except that the contigs
            from the 1.5x poodle assembly (Kirkness 2003) were used instead of
            WGS reads. Since this
            sequence lacked base quality scores, arbitrary quality scores of
            phred 40 were assigned
            before the poddle seqeunce was placed uniquely on the CanFam1.0
            boxer assembly and SNP
            detection was carried out by SSAHA-SNP. 1637780 SNPs were annotated
            with alleles from the
            boxer and the poodle. The validation rate for these SNPs was
            estimated at approximately TBD%.
            Internal-WGA-discovery (I-WGA):
            A third set of SNPs were discovered by comparing reads in the WGA
            assembly. SNPs were
            defined as mismatch positions that had a base quality of >= 30 on
            both reads in a region
            that aligned without gaps, and with at most one additional mismatch
            in the ten flanking
            bases. For each allele, at least one additional read had to confirm
            it. 731476 SNPs were
            annotated with alleles between the two boxer alleles. The
            validation rate for these SNPs
            was estimated at approximately TBD%.
FEATURES             Location/Qualifiers
     source          1. .478
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /strain="Beagle"
                     /db_xref="taxon:9615"
                     /map="+ 32 22-427 4685905-4685500"
                     /clone_lib="Beagle"
     STS             <1. .>478
ORIGIN

Alignment Scores:
Pred. No.:                 115          Length:          478
Score:                     42.00        Matches:         6
Percent Similarity:        100.00%      Conservative:    0
Best Local Similarity:     100.00%      Mismatches:      0
Query Match:               100.00%      Indels:          0
DB:                        10           Gaps:            0

US-09-858-332G-28 (1-6) x BV281117 (1-478)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        212 CACAACCACAACCACAAC 229

RESULT 86
BTU50559/c
LOCUS       BTU50559                480 bp    mRNA    linear   MAM 27-MAR-2002
DEFINITION  Bos taurus clone F1 IgM mRNA, partial cds.
ACCESSION   U50559
VERSION     U50559.1  GI:1399522
KEYWORDS    .
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 480)
  AUTHORS   Berens,S.J., Wylie,D. and Lopez,O.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-1996) Osvaldo J. Lopez, Biological Sciences,
            University of Nebraska, 325 Manter Hall, Lincoln, NE 68588-0118,
            USA
REFERENCE   2  (bases 1 to 480)
  AUTHORS   Berens,S.J., Wylie,D.E. and Lopez,O.J.
  TITLE     Use of a single VH family and long CDR3s in the variable region of
            cattle Ig heavy chains
  JOURNAL   Int. Immunol. 9 (1), 189-199 (1997)
  PUBMED    9043960
FEATURES             Location/Qualifiers
     source          1. .480
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /clone="F1"
                     /tissue_type="spleen"
                     /dev_stage="150 day fetus"
     CDS             <1. .>480
                     /note="rearranged VDJ exon; immunoglobulin"
                     /codon_start=1
                     /product="IgM"
                     /protein_id="AAB03276.1"
                     /db_xref="GI:1399523"
                     /translation="WRLVGVVTLICSKMNPLWTLLFVLSAPRGVLSQVQLRESGPSLV
                     KPSQTLSLTCTASGFSLSDKAVGWVRQAPGKALEWLGGIDTGGSTGYNPGLKSRLSIT
                     KDNSKSQVSLSVSSVTTEDSATYYCTTVHRSCYSGYGYGYGCGYGDAWGQGLLVTVSS
                     "
ORIGIN

Alignment Scores:
Pred. No.:                 115          Length:          480
Score:                     42.00        Matches:         6
Percent Similarity:        100.00%      Conservative:    0
Best Local Similarity:     100.00%      Mismatches:      0
Query Match:               100.00%      Indels:          0
```

```
DB:                    4              Gaps:          0

US-09-858-332G-28 (1-6) x BTU50559 (1-480)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        439 CATAACCACAACCATAAC 422

RESULT 87
AY768786/c
LOCUS       AY768786                480 bp    DNA     linear   PLN 17-MAY-2005
DEFINITION  Passiflora edulis f. flavicarpa microsatellite PE08 sequence.
ACCESSION   AY768786
VERSION     AY768786.1  GI:54301286
KEYWORDS    .
SOURCE      Passiflora edulis f. flavicarpa
  ORGANISM  Passiflora edulis f. flavicarpa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Malpighiales; Passifloraceae; Passiflora.
REFERENCE   1  (bases 1 to 480)
  AUTHORS   Oliveira,E.J., Padua,J.G., Zucchi,M.I., Camargo,L.E., Fungaro,M.H.
            and Vieira,M.L.
  TITLE     Development and characterization of microsatellite markers from the
            yellow passion fruit (Passiflora edulis f. flavicarpa)
  JOURNAL   Mol. Ecol. Notes 5 (2), 331-333 (2005)
REFERENCE   2  (bases 1 to 480)
  AUTHORS   Oliveira,E.J., Padua,J.G., Zucchi,M.I., Camargo,L.E.A. and
            Vieira,M.L.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-SEP-2004) Genetica, ESALQ/USP, Av. Padua Dias 11,
            Piracicaba, SP 13418900, Brazil
FEATURES             Location/Qualifiers
     source          1. .480
                     /organism="Passiflora edulis f. flavicarpa"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:237848"
                     /note="authority: Passiflora edulis Sims. f. flavicarpa
                     Deg; forma: flavicarpa"
     repeat_region   1. .480
                     /note="microsatellite PE08"
                     /rpt_type=tandem
ORIGIN

Alignment Scores:
Pred. No.:              115           Length:        480
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     15            Gaps:          0

US-09-858-332G-28 (1-6) x AY768786 (1-480)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        232 CACAACCACAACCACAAC 215

RESULT 88
AF097203/c
LOCUS       AF097203                489 bp    mRNA    linear   MAM 17-DEC-2003
DEFINITION  Bos taurus clone BF1H1 immunoglobulin heavy chain variable region
            mRNA, partial cds.
ACCESSION   AF097203
VERSION     AF097203.1  GI:3834646
KEYWORDS    .
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 489)
  AUTHORS   Saini,S.S. and Kaushik,A.
  TITLE     Extensive CDR3H length heterogeneity exists in bovine foetal VDJ
            rearrangements
  JOURNAL   Scand. J. Immunol. 55 (2), 140-148 (2002)
  PUBMED    11896930
REFERENCE   2  (bases 1 to 489)
  AUTHORS   Saini,S.S., Allore,B., Jacobs,R. and Kaushik,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1998) Pathobiology, University of Guelph, Guelph,
            ON N1G 2W1, Canada
FEATURES             Location/Qualifiers
     source          1. .489
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /clone="BF1H1"
                     /cell_type="mouse x Bovine heterohybridoma"
                     /dev_stage="fetal"
                     /note="PCR product"
     CDS             <1. .>489
                     /note="VDJ region"
                     /codon_start=1
                     /product="immunoglobulin heavy chain variable region"
                     /protein_id="AAC71038.1"
                     /db_xref="GI:3834647"
                     /translation="QVQLRESGPSLVKPSQTLSLTCTASGFSLSDKAVGWVRQAPGKA
                     LEWLGGIDTGGSTGYNPGLKSRLSITKDNSKSQVSLSVSSVTTEDSATYYCTTVHPSP
                     DGYSYGYGCGYGYGCSGYDCYGYGGYGYGGYGGYSSYSYSYSPRKYYVDAWGQGLLVT
                     VSS"
ORIGIN

Alignment Scores:
Pred. No.:              117           Length:        489
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     4             Gaps:          0

US-09-858-332G-28 (1-6) x AF097203 (1-489)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        346 CATAACCATAACCACAAC 329

RESULT 89
AF097204/c
LOCUS       AF097204                492 bp    mRNA    linear   MAM 17-DEC-2003
DEFINITION  Bos taurus clone BF4E9 immunoglobulin heavy chain variable region
            mRNA, partial cds.
ACCESSION   AF097204
VERSION     AF097204.1  GI:3834648
KEYWORDS    .
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 492)
  AUTHORS   Saini,S.S. and Kaushik,A.
  TITLE     Extensive CDR3H length heterogeneity exists in bovine foetal VDJ
            rearrangements
  JOURNAL   Scand. J. Immunol. 55 (2), 140-148 (2002)
  PUBMED    11896930
REFERENCE   2  (bases 1 to 492)
  AUTHORS   Saini,S.S., Allore,B., Jacobs,R. and Kaushik,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1998) Pathobiology, University of Guelph, Guelph,
            ON N1G 2W1, Canada
FEATURES             Location/Qualifiers
     source          1. .492
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
```

```
                     /clone="BF4E9"
                     /cell_type="mouse x Bovine heterohybridoma"
                     /dev_stage="fetal"
                     /note="PCR product"
     CDS             <1. .>492
                     /note="VDJ region"
                     /codon_start=1
                     /product="immunoglobulin heavy chain variable region"
                     /protein_id="AAC71039.1"
                     /db_xref="GI:3834649"
                     /translation="QVQLRESGPSLVKPSQTLSLTCTASGFSLSDKAVGWVRQAPGKA
                     LEWLGGIDTGGSTGYNPGLKSRLSITKDNSKSQVSLSVSSVTTEDSATYYCTTVHQIF
                     CPDGYSYGYGCGYGYGCSGYDCYGYGGYGYGGYGGYSSYSYSYSYEYYGDAWGQGLLV
                     TVSS"
ORIGIN

Alignment Scores:
Pred. No.:                 118           Length:          492
Score:                     42.00         Matches:         6
Percent Similarity:        100.00%       Conservative:    0
Best Local Similarity:     100.00%       Mismatches:      0
Query Match:               100.00%       Indels:          0
DB:                        4             Gaps:            0

US-09-858-332G-28 (1-6) x AF097204 (1-492)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        352 CATAACCATAACCACAAC 335

RESULT 90
AF187065
LOCUS       AF187065                519 bp    mRNA    linear   ROD 11-JUN-2000
DEFINITION  Rattus norvegicus p75NTR-associated cell death executor (Nade)
            mRNA, complete cds.
ACCESSION   AF187065
VERSION     AF187065.1  GI:8452895
KEYWORDS    .
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 519)
  AUTHORS   Mukai,J., Hachiya,T., Shoji-Hoshino,S., Kimura,M.T., Nadano,D.,
            Suvanto,P., Hanaoka,T., Li,Y., Irie,S., Greene,L.A. and Sato,T.A.
  TITLE     NADE, a p75NTR-associated cell death executor, is involved in
            signal transduction mediated by the common neurotrophin receptor
            p75NTR
  JOURNAL   J. Biol. Chem. 275 (23), 17566-17570 (2000)
   PUBMED   10764727
REFERENCE   2  (bases 1 to 519)
  AUTHORS   Mukai,J., Hachiya,T., Hoshino,S., Kimura,M., Nadano,D., Suvanto,P.,
            Hanaoka,T., Li,Y., Irie,S. and Sato,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-1999) Otolaryngology/Pathology, Columbia
            University, 630 West 168th St., P&S 11-451, New York, NY 10032, USA
FEATURES             Location/Qualifiers
     source          1. .519
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
     gene            1. .519
                     /gene="Nade"
     CDS             44. .406
                     /gene="Nade"
                     /note="involved in the common neurotrophin receptor
                     p75NTR-mediated signal transduction; NADE"
                     /codon_start=1
                     /product="p75NTR-associated cell death executor"
                     /protein_id="AAF75130.1"
                     /db_xref="GI:8452896"
                     /translation="MEQPLQNGQEDRPVGGGEGHQPAANNNNHNHNHNHNHNHNHNH
                     RRGQARRLAPNFRWAIPNRQMNDGLGGDGDDMEMFMEEMREIRRKLRELQLRNCLRIL
                     MGELSNHHDHHDEFCLMP"
ORIGIN

Alignment Scores:
Pred. No.:                 124           Length:          519
Score:                     42.00         Matches:         6
Percent Similarity:        100.00%       Conservative:    0
Best Local Similarity:     100.00%       Mismatches:      0
Query Match:               100.00%       Indels:          0
DB:                        9             Gaps:            0

US-09-858-332G-28 (1-6) x AF187065 (1-519)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        128 CACAACCATAACCACAAC 145

RESULT 91
AR554223
LOCUS       AR554223                528 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 9354 from patent US 6747137.
ACCESSION   AR554223
VERSION     AR554223.1  GI:53947398
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 528)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 9354 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .528
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:                 126           Length:          528
Score:                     42.00         Matches:         6
Percent Similarity:        100.00%       Conservative:    0
Best Local Similarity:     100.00%       Mismatches:      0
Query Match:               100.00%       Indels:          0
DB:                        6             Gaps:            0

US-09-858-332G-28 (1-6) x AR554223 (1-528)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        244 CATAATCATAATCATAAT 261

RESULT 92
AY677214
LOCUS       AY677214                530 bp    DNA     linear   PLN 17-MAR-2005
DEFINITION  Campylocarpon pseudofasciculare strain CBS 112679 beta-tublulin
            gene, partial cds.
ACCESSION   AY677214
VERSION     AY677214.1  GI:57901548
KEYWORDS    .
SOURCE      Campylocarpon pseudofasciculare
  ORGANISM  Campylocarpon pseudofasciculare
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Hypocreomycetidae; Hypocreales; Nectriaceae; mitosporic
            Nectriaceae; Campylocarpon.
REFERENCE   1  (bases 1 to 530)
  AUTHORS   Halleen,F., Schroers,H.-J., Groenewald,J.Z. and Crous,P.W.
  TITLE     Novel species of Cylindrocarpon (Neonectria) and Campylocarpon gen.
            nov. associated with black foot disease of grapevines (Vitis spp.)
  JOURNAL   Stud. Mycol. 50, 431-455 (2004)
```

```
REFERENCE   2  (bases 1 to 530)
  AUTHORS   Halleen,F., Schroers,H.-J., Groenewald,J.Z. and Crous,P.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-2004) Fungal Biodiversity Centre CBS, Uppsalalaan
            8, Utrecht 3584 CT, The Netherlands
FEATURES             Location/Qualifiers
     source          1. .530
                     /organism="Campylocarpon pseudofasciculare"
                     /mol_type="genomic DNA"
                     /strain="CBS 112679"
                     /db_xref="taxon:307969"
     mRNA            join(<148. .171,233. .355,421. .>530)
                     /product="beta-tublulin"
     CDS             join(<148. .171,233. .355,421. .>530)
                     /codon_start=1
                     /product="beta-tublulin"
                     /protein_id="AAW57943.1"
                     /db_xref="GI:57901549"
                     /translation="VHLQTGQCGNQIGAAFWQTISGEHGLDSNGVYNGTSELQLERMS
                     VYFNEASGNKYVPRAVLVDLEPGTMDAVRAGPFGQLFRPDN"
ORIGIN

Alignment Scores:
Pred. No.:              127             Length:         530
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x AY677214 (1-530)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         54 CACAACCACAACCACAAC 71

RESULT 93
AY677216
LOCUS       AY677216                 536 bp    DNA     linear   PLN 17-MAR-2005
DEFINITION  Campylocarpon fasciculare strain CBS 112612 beta-tublulin gene,
            partial cds.
ACCESSION   AY677216
VERSION     AY677216.1  GI:57901552
KEYWORDS    .
SOURCE      Campylocarpon fasciculare
  ORGANISM  Campylocarpon fasciculare
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Hypocreomycetidae; Hypocreales; Nectriaceae; mitosporic
            Nectriaceae; Campylocarpon.
REFERENCE   1  (bases 1 to 536)
  AUTHORS   Halleen,F., Schroers,H.-J., Groenewald,J.Z. and Crous,P.W.
  TITLE     Novel species of Cylindrocarpon (Neonectria) and Campylocarpon gen.
            nov. associated with black foot disease of grapevines (Vitis spp.)
  JOURNAL   Stud. Mycol. 50, 431-455 (2004)
REFERENCE   2  (bases 1 to 536)
  AUTHORS   Halleen,F., Schroers,H.-J., Groenewald,J.Z. and Crous,P.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-2004) Fungal Biodiversity Centre CBS, Uppsalalaan
            8, Utrecht 3584 CT, The Netherlands
FEATURES             Location/Qualifiers
     source          1. .536
                     /organism="Campylocarpon fasciculare"
                     /mol_type="genomic DNA"
                     /strain="CBS 112612"
                     /db_xref="taxon:307939"
     mRNA            join(<153. .176,238. .360,427. .>536)
                     /product="beta-tublulin"
     CDS             join(<153. .176,238. .360,427. .>536)
                     /codon_start=1
                     /product="beta-tublulin"
                     /protein_id="AAW57945.1"
                     /db_xref="GI:57901553"
                     /translation="VHLQTGQCGNQIGAAFWQTISGEHGLDSNGVYNGTSELQLERMS
                     VYFNEASGNKYVPRAVLVDLEPGTMDAVRAGPFGQLFRPDN"
ORIGIN

Alignment Scores:
Pred. No.:              128             Length:         536
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x AY677216 (1-536)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         47 CACAACCACAACCACAAC 64

RESULT 94
AY677217
LOCUS       AY677217                 536 bp    DNA     linear   PLN 17-MAR-2005
DEFINITION  Campylocarpon fasciculare strain CBS 113560 beta-tublulin gene,
            partial cds.
ACCESSION   AY677217
VERSION     AY677217.1  GI:57901554
KEYWORDS    .
SOURCE      Campylocarpon fasciculare
  ORGANISM  Campylocarpon fasciculare
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Hypocreomycetidae; Hypocreales; Nectriaceae; mitosporic
            Nectriaceae; Campylocarpon.
REFERENCE   1  (bases 1 to 536)
  AUTHORS   Halleen,F., Schroers,H.-J., Groenewald,J.Z. and Crous,P.W.
  TITLE     Novel species of Cylindrocarpon (Neonectria) and Campylocarpon gen.
            nov. associated with black foot disease of grapevines (Vitis spp.)
  JOURNAL   Stud. Mycol. 50, 431-455 (2004)
REFERENCE   2  (bases 1 to 536)
  AUTHORS   Halleen,F., Schroers,H.-J., Groenewald,J.Z. and Crous,P.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-2004) Fungal Biodiversity Centre CBS, Uppsalalaan
            8, Utrecht 3584 CT, The Netherlands
FEATURES             Location/Qualifiers
     source          1. .536
                     /organism="Campylocarpon fasciculare"
                     /mol_type="genomic DNA"
                     /strain="CBS 113560"
                     /db_xref="taxon:307939"
     mRNA            join(<153. .176,238. .360,427. .>536)
                     /product="beta-tublulin"
     CDS             join(<153. .176,238. .360,427. .>536)
                     /codon_start=1
                     /product="beta-tublulin"
                     /protein_id="AAW57946.1"
                     /db_xref="GI:57901555"
                     /translation="VHLQTGQCGNQIGAAFWQTISGEHGLDSNGVYNGTSELQLERMS
                     VYFNEASGNKYVPRAVLVDLEPGTMDAVRAGPFGQLFRPDN"
ORIGIN

Alignment Scores:
Pred. No.:              128             Length:         536
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x AY677217 (1-536)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         47 CACAACCACAACCACAAC 64

RESULT 95
AY677218
```

```
LOCUS       AY677218                 536 bp    DNA     linear   PLN 17-MAR-2005
DEFINITION  Campylocarpon fasciculare strain CBS 113559 beta-tublulin gene,
            partial cds.
ACCESSION   AY677218
VERSION     AY677218.1  GI:57901556
KEYWORDS    .
SOURCE      Campylocarpon fasciculare
  ORGANISM  Campylocarpon fasciculare
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Hypocreomycetidae; Hypocreales; Nectriaceae; mitosporic
            Nectriaceae; Campylocarpon.
REFERENCE   1  (bases 1 to 536)
  AUTHORS   Halleen,F., Schroers,H.-J., Groenewald,J.Z. and Crous,P.W.
  TITLE     Novel species of Cylindrocarpon (Neonectria) and Campylocarpon gen.
            nov. associated with black foot disease of grapevines (Vitis spp.)
  JOURNAL   Stud. Mycol. 50, 431-455 (2004)
REFERENCE   2  (bases 1 to 536)
  AUTHORS   Halleen,F., Schroers,H.-J., Groenewald,J.Z. and Crous,P.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-2004) Fungal Biodiversity Centre CBS, Uppsalalaan
            8, Utrecht 3584 CT, The Netherlands
FEATURES             Location/Qualifiers
     source          1. .536
                     /organism="Campylocarpon fasciculare"
                     /mol_type="genomic DNA"
                     /strain="CBS 113559"
                     /db_xref="taxon:307939"
     mRNA            join(<153. .176,238. .360,427. .>536)
                     /product="beta-tublulin"
     CDS             join(<153. .176,238. .360,427. .>536)
                     /codon_start=1
                     /product="beta-tublulin"
                     /protein_id="AAW57947.1"
                     /db_xref="GI:57901557"
                     /translation="VHLQTGQCGNQIGAAFWQTISGEHGLDSNGVYNGTSELQLERMS
                     VYFNEASGNKYVPRAVLVDLEPGTMDAVRAGPFGQLFRPDN"
ORIGIN

Alignment Scores:
Pred. No.:                128             Length:          536
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       15              Gaps:            0

US-09-858-332G-28 (1-6) x AY677218 (1-536)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         47 CACAACCACAACCACAAC 64

RESULT 96
AY677219
LOCUS       AY677219                 536 bp    DNA     linear   PLN 17-MAR-2005
DEFINITION  Campylocarpon fasciculare strain CBS 112600 beta-tublulin gene,
            partial cds.
ACCESSION   AY677219
VERSION     AY677219.1  GI:57901558
KEYWORDS    .
SOURCE      Campylocarpon fasciculare
  ORGANISM  Campylocarpon fasciculare
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Hypocreomycetidae; Hypocreales; Nectriaceae; mitosporic
            Nectriaceae; Campylocarpon.
REFERENCE   1  (bases 1 to 536)
  AUTHORS   Halleen,F., Schroers,H.-J., Groenewald,J.Z. and Crous,P.W.
  TITLE     Novel species of Cylindrocarpon (Neonectria) and Campylocarpon gen.
            nov. associated with black foot disease of grapevines (Vitis spp.)
  JOURNAL   Stud. Mycol. 50, 431-455 (2004)
REFERENCE   2  (bases 1 to 536)
  AUTHORS   Halleen,F., Schroers,H.-J., Groenewald,J.Z. and Crous,P.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-2004) Fungal Biodiversity Centre CBS, Uppsalalaan
            8, Utrecht 3584 CT, The Netherlands
FEATURES             Location/Qualifiers
     source          1. .536
                     /organism="Campylocarpon fasciculare"
                     /mol_type="genomic DNA"
                     /strain="CBS 112600"
                     /db_xref="taxon:307939"
     mRNA            join(<153. .176,238. .360,427. .>536)
                     /product="beta-tublulin"
     CDS             join(<153. .176,238. .360,427. .>536)
                     /codon_start=1
                     /product="beta-tublulin"
                     /protein_id="AAW57948.1"
                     /db_xref="GI:57901559"
                     /translation="VHLQTGQCGNQIGAAFWQTISGEHGLDSNGVYNGTSELQLERMS
                     VYFNEASGNKYVPRAVLVDLEPGTMDAVRAGPFGQLFRPDN"
ORIGIN

Alignment Scores:
Pred. No.:                128             Length:          536
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       15              Gaps:            0

US-09-858-332G-28 (1-6) x AY677219 (1-536)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         47 CACAACCACAACCACAAC 64

RESULT 97
AY677220
LOCUS       AY677220                 536 bp    DNA     linear   PLN 17-MAR-2005
DEFINITION  Campylocarpon fasciculare strain CBS 112614 beta-tublulin gene,
            partial cds.
ACCESSION   AY677220
VERSION     AY677220.1  GI:57901560
KEYWORDS    .
SOURCE      Campylocarpon fasciculare
  ORGANISM  Campylocarpon fasciculare
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Hypocreomycetidae; Hypocreales; Nectriaceae; mitosporic
            Nectriaceae; Campylocarpon.
REFERENCE   1  (bases 1 to 536)
  AUTHORS   Halleen,F., Schroers,H.-J., Groenewald,J.Z. and Crous,P.W.
  TITLE     Novel species of Cylindrocarpon (Neonectria) and Campylocarpon gen.
            nov. associated with black foot disease of grapevines (Vitis spp.)
  JOURNAL   Stud. Mycol. 50, 431-455 (2004)
REFERENCE   2  (bases 1 to 536)
  AUTHORS   Halleen,F., Schroers,H.-J., Groenewald,J.Z. and Crous,P.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-2004) Fungal Biodiversity Centre CBS, Uppsalalaan
            8, Utrecht 3584 CT, The Netherlands
FEATURES             Location/Qualifiers
     source          1. .536
                     /organism="Campylocarpon fasciculare"
                     /mol_type="genomic DNA"
                     /strain="CBS 112614"
                     /db_xref="taxon:307939"
     mRNA            join(<153. .176,238. .360,427. .>536)
                     /product="beta-tublulin"
     CDS             join(<153. .176,238. .360,427. .>536)
                     /codon_start=1
                     /product="beta-tublulin"
                     /protein_id="AAW57949.1"
                     /db_xref="GI:57901561"
                     /translation="VHLQTGQCGNQIGAAFWQTISGEHGLDSNGVYNGTSELQLERMS
                     VYFNEASGNKYVPRAVLVDLEPGTMDAVRAGPFGQLFRPDN"
ORIGIN
```

```
Alignment Scores:
Pred. No.:                128             Length:         536
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       15              Gaps:           0

US-09-858-332G-28 (1-6) x AY677220 (1-536)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         47 CACAACCACAACCACAAC 64

RESULT 98
AY677221
LOCUS       AY677221                  536 bp    DNA     linear   PLN 17-MAR-2005
DEFINITION  Campylocarpon fasciculare strain CBS 112613 beta-tublulin gene,
            partial cds.
ACCESSION   AY677221
VERSION     AY677221.1  GI:57901562
KEYWORDS    .
SOURCE      Campylocarpon fasciculare
  ORGANISM  Campylocarpon fasciculare
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Hypocreomycetidae; Hypocreales; Nectriaceae; mitosporic
            Nectriaceae; Campylocarpon.
REFERENCE   1  (bases 1 to 536)
  AUTHORS   Halleen,F., Schroers,H.-J., Groenewald,J.Z. and Crous,P.W.
  TITLE     Novel species of Cylindrocarpon (Neonectria) and Campylocarpon gen.
            nov. associated with black foot disease of grapevines (Vitis spp.)
  JOURNAL   Stud. Mycol. 50, 431-455 (2004)
REFERENCE   2  (bases 1 to 536)
  AUTHORS   Halleen,F., Schroers,H.-J., Groenewald,J.Z. and Crous,P.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-2004) Fungal Biodiversity Centre CBS, Uppsalalaan
            8, Utrecht 3584 CT, The Netherlands
FEATURES             Location/Qualifiers
     source          1. .536
                     /organism="Campylocarpon fasciculare"
                     /mol_type="genomic DNA"
                     /strain="CBS 112613"
                     /db_xref="taxon:307939"
     mRNA            join(<153. .176,238. .360,427. .>536)
                     /product="beta-tublulin"
     CDS             join(<153. .176,238. .360,427. .>536)
                     /codon_start=1
                     /product="beta-tublulin"
                     /protein_id="AAW57950.1"
                     /db_xref="GI:57901563"
                     /translation="VHLQTGQCGNQIGAAFWQTISGEHGLDSNGVYNGTSELQLERMS
                     VYFNEASGNKYVPRAVLVDLEPGTMDAVRAGPFGQLFRPDN"
ORIGIN

Alignment Scores:
Pred. No.:                128             Length:         536
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       15              Gaps:           0

US-09-858-332G-28 (1-6) x AY677221 (1-536)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         47 CACAACCACAACCACAAC 64

RESULT 99
A57711
LOCUS       A57711                    537 bp    DNA     linear   PAT 03-MAR-1998
DEFINITION  Sequence 21 from Patent WO9634103.
ACCESSION   A57711
VERSION     A57711.1  GI:3713535
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified sequences.
REFERENCE   1
  AUTHORS   Hamers,R. and Muyldermans,S.
  TITLE     Variable fragments of immunoglobulins - use for therapeutic or
            veterinary purposes
  JOURNAL   Patent: WO 9634103-A 21 31-OCT-1996;
            UNIV BRUXELLES (BE)
COMMENT     Other publication AU 5647896 961118.
FEATURES             Location/Qualifiers
     source          1. .537
                     /organism="unidentified"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32644"
     CDS             1. .537
                     /note="unnamed protein product; Protein sequence is in
                     conflict with the conceptual translation"
                     /codon_start=1
                     /protein_id="CAA03459.1"
                     /db_xref="GI:3713536"
                     /translation="DVQLQASGGGSVQAGGSLRLSCAASGYTIGPYCMGWFRQAPGKE
                     REGVAAINMGGGITYYADSVKGRFTISQDNAKNTVYLLMNSLEPEDTAIYYCAADSTI
                     YASYYECGHGLSTGGYGYDSWGQGTQVTVSSEPKIPQPQPKPQPQPQPKPQPKPEP
                     MECGRYPYDVPDYGSGRA"
ORIGIN

Alignment Scores:
Pred. No.:                129             Length:         537
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x A57711 (1-537)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        431 CACAACCACAACCACAAC 448

RESULT 100
AX813676
LOCUS       AX813676                  549 bp    DNA     linear   PAT 02-DEC-2003
DEFINITION  Sequence 9 from Patent WO03062415.
ACCESSION   AX813676
VERSION     AX813676.1  GI:38636044
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Drabek,D., Dekker,S. and Grosveld,F.G.
  TITLE     Transgenic animal and methods
  JOURNAL   Patent: WO 03062415-A 9 31-JUL-2003;
            Erasmus Unviersiteit (NL)
FEATURES             Location/Qualifiers
     source          1. .549
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Anti p15 A5"
ORIGIN

Alignment Scores:
Pred. No.:                131             Length:         549
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0
```

```
US-09-858-332G-28 (1-6) x AX813676 (1-549)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        449 CACAACCACAACCACAAC 466

RESULT 101
AX813677
LOCUS       AX813677                549 bp    DNA     linear   PAT 02-DEC-2003
DEFINITION  Sequence 10 from Patent WO03062415.
ACCESSION   AX813677
VERSION     AX813677.1  GI:38636045
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Drabek,D., Dekker,S. and Grosveld,F.G.
  TITLE     Transgenic animal and methods
  JOURNAL   Patent: WO 03062415-A 10 31-JUL-2003;
            Erasmus Unviersiteit (NL)
FEATURES             Location/Qualifiers
     source          1. .549
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="anti p15 E11"
ORIGIN

Alignment Scores:
Pred. No.:              131           Length:         549
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     6             Gaps:           0

US-09-858-332G-28 (1-6) x AX813677 (1-549)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        449 CACAACCACAACCACAAC 466

RESULT 102
AX813679
LOCUS       AX813679                549 bp    DNA     linear   PAT 02-DEC-2003
DEFINITION  Sequence 12 from Patent WO03062415.
ACCESSION   AX813679
VERSION     AX813679.1  GI:38636047
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Drabek,D., Dekker,S. and Grosveld,F.G.
  TITLE     Transgenic animal and methods
  JOURNAL   Patent: WO 03062415-A 12 31-JUL-2003;
            Erasmus Unviersiteit (NL)
FEATURES             Location/Qualifiers
     source          1. .549
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="anti p15 B10"
ORIGIN

Alignment Scores:
Pred. No.:              131           Length:         549
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     6             Gaps:           0

US-09-858-332G-28 (1-6) x AX813679 (1-549)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        449 CACAACCACAACCACAAC 466

RESULT 103
AY677222
LOCUS       AY677222                553 bp    DNA     linear   PLN 17-MAR-2005
DEFINITION  Campylocarpon fasciculare strain CBS 113557 beta-tublulin gene,
            partial cds.
ACCESSION   AY677222
VERSION     AY677222.1  GI:57901564
KEYWORDS    .
SOURCE      Campylocarpon fasciculare
  ORGANISM  Campylocarpon fasciculare
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Hypocreomycetidae; Hypocreales; Nectriaceae; mitosporic
            Nectriaceae; Campylocarpon.
REFERENCE   1  (bases 1 to 553)
  AUTHORS   Halleen,F., Schroers,H.-J., Groenewald,J.Z. and Crous,P.W.
  TITLE     Novel species of Cylindrocarpon (Neonectria) and Campylocarpon gen.
            nov. associated with black foot disease of grapevines (Vitis spp.)
  JOURNAL   Stud. Mycol. 50, 431-455 (2004)
REFERENCE   2  (bases 1 to 553)
  AUTHORS   Halleen,F., Schroers,H.-J., Groenewald,J.Z. and Crous,P.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-2004) Fungal Biodiversity Centre CBS, Uppsalalaan
            8, Utrecht 3584 CT, The Netherlands
FEATURES             Location/Qualifiers
     source          1. .553
                     /organism="Campylocarpon fasciculare"
                     /mol_type="genomic DNA"
                     /strain="CBS 113557"
                     /db_xref="taxon:307939"
     mRNA            join(<153. .176,238. .360,427. .>537)
                     /product="beta-tublulin"
     CDS             join(<153. .176,238. .360,427. .>537)
                     /codon_start=1
                     /product="beta-tublulin"
                     /protein_id="AAW57951.1"
                     /db_xref="GI:57901565"
                     /translation="VHLQTGQCGNQIGAAFWQTISGEHGLDSNGVYNGTSELQLERMS
                     VYFNEASGNKYVPRAVLVDLEPGTMDAVRAGPFGQLFRPDNF"
ORIGIN

Alignment Scores:
Pred. No.:              132           Length:         553
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     15            Gaps:           0

US-09-858-332G-28 (1-6) x AY677222 (1-553)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         47 CACAACCACAACCACAAC 64

RESULT 104
AY677223
LOCUS       AY677223                553 bp    DNA     linear   PLN 17-MAR-2005
DEFINITION  Campylocarpon fasciculare strain CBS 113554 beta-tublulin gene,
            partial cds.
ACCESSION   AY677223
VERSION     AY677223.1  GI:57901566
KEYWORDS    .
SOURCE      Campylocarpon fasciculare
  ORGANISM  Campylocarpon fasciculare
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
```

```
            Hypocreomycetidae; Hypocreales; Nectriaceae; mitosporic
            Nectriaceae; Campylocarpon.
REFERENCE   1  (bases 1 to 553)
  AUTHORS   Halleen,F., Schroers,H.-J., Groenewald,J.Z. and Crous,P.W.
  TITLE     Novel species of Cylindrocarpon (Neonectria) and Campylocarpon gen.
            nov. associated with black foot disease of grapevines (Vitis spp.)
  JOURNAL   Stud. Mycol. 50, 431-455 (2004)
REFERENCE   2  (bases 1 to 553)
  AUTHORS   Halleen,F., Schroers,H.-J., Groenewald,J.Z. and Crous,P.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-2004) Fungal Biodiversity Centre CBS, Uppsalalaan
            8, Utrecht 3584 CT, The Netherlands
FEATURES             Location/Qualifiers
     source          1. .553
                     /organism="Campylocarpon fasciculare"
                     /mol_type="genomic DNA"
                     /strain="CBS 113554"
                     /db_xref="taxon:307939"
     mRNA            join(<153. .176,238. .360,427. .>537)
                     /product="beta-tublulin"
     CDS             join(<153. .176,238. .360,427. .>537)
                     /codon_start=1
                     /product="beta-tublulin"
                     /protein_id="AAW57952.1"
                     /db_xref="GI:57901567"
                     /translation="VHLQTGQCGNQIGAAFWQTISGEHGLDSNGVYNGTSELQLERMS
                     VYFNEASGNKYVPRAVLVDLEPGTMDAVRAGPFGQLFRPDNF"
ORIGIN

Alignment Scores:
Pred. No.:                132           Length:         553
Score:                    42.00         Matches:        6
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              100.00%       Indels:         0
DB:                       15            Gaps:           0

US-09-858-332G-28 (1-6) x AY677223 (1-553)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         47 CACAACCACAACCACAAC 64

RESULT 105
AY677224
LOCUS       AY677224                553 bp    DNA     linear   PLN 17-MAR-2005
DEFINITION  Campylocarpon fasciculare strain CBS 113558 beta-tublulin gene,
            partial cds.
ACCESSION   AY677224
VERSION     AY677224.1  GI:57901568
KEYWORDS    .
SOURCE      Campylocarpon fasciculare
  ORGANISM  Campylocarpon fasciculare
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Hypocreomycetidae; Hypocreales; Nectriaceae; mitosporic
            Nectriaceae; Campylocarpon.
REFERENCE   1  (bases 1 to 553)
  AUTHORS   Halleen,F., Schroers,H.-J., Groenewald,J.Z. and Crous,P.W.
  TITLE     Novel species of Cylindrocarpon (Neonectria) and Campylocarpon gen.
            nov. associated with black foot disease of grapevines (Vitis spp.)
  JOURNAL   Stud. Mycol. 50, 431-455 (2004)
REFERENCE   2  (bases 1 to 553)
  AUTHORS   Halleen,F., Schroers,H.-J., Groenewald,J.Z. and Crous,P.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-2004) Fungal Biodiversity Centre CBS, Uppsalalaan
            8, Utrecht 3584 CT, The Netherlands
FEATURES             Location/Qualifiers
     source          1. .553
                     /organism="Campylocarpon fasciculare"
                     /mol_type="genomic DNA"
                     /strain="CBS 113558"
                     /db_xref="taxon:307939"
     mRNA            join(<153. .176,238. .360,427. .>537)
                     /product="beta-tublulin"
     CDS             join(<153. .176,238. .360,427. .>537)
                     /codon_start=1
                     /product="beta-tublulin"
                     /protein_id="AAW57953.1"
                     /db_xref="GI:57901569"
                     /translation="VHLQTGQCGNQIGAAFWQTISGEHGLDSNGVYNGTSELQLERMS
                     VYFNEASGNKYVPRAVLVDLEPGTMDAVRAGPFGQLFRPDNF"
ORIGIN

Alignment Scores:
Pred. No.:                132           Length:         553
Score:                    42.00         Matches:        6
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              100.00%       Indels:         0
DB:                       15            Gaps:           0

US-09-858-332G-28 (1-6) x AY677224 (1-553)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         47 CACAACCACAACCACAAC 64

RESULT 106
AY677225
LOCUS       AY677225                557 bp    DNA     linear   PLN 17-MAR-2005
DEFINITION  Campylocarpon fasciculare strain CBS 112611 beta-tublulin gene,
            partial cds.
ACCESSION   AY677225
VERSION     AY677225.1  GI:57901570
KEYWORDS    .
SOURCE      Campylocarpon fasciculare
  ORGANISM  Campylocarpon fasciculare
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Hypocreomycetidae; Hypocreales; Nectriaceae; mitosporic
            Nectriaceae; Campylocarpon.
REFERENCE   1  (bases 1 to 557)
  AUTHORS   Halleen,F., Schroers,H.-J., Groenewald,J.Z. and Crous,P.W.
  TITLE     Novel species of Cylindrocarpon (Neonectria) and Campylocarpon gen.
            nov. associated with black foot disease of grapevines (Vitis spp.)
  JOURNAL   Stud. Mycol. 50, 431-455 (2004)
REFERENCE   2  (bases 1 to 557)
  AUTHORS   Halleen,F., Schroers,H.-J., Groenewald,J.Z. and Crous,P.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-2004) Fungal Biodiversity Centre CBS, Uppsalalaan
            8, Utrecht 3584 CT, The Netherlands
FEATURES             Location/Qualifiers
     source          1. .557
                     /organism="Campylocarpon fasciculare"
                     /mol_type="genomic DNA"
                     /strain="CBS 112611"
                     /db_xref="taxon:307939"
     mRNA            join(<153. .176,238. .360,427. .>537)
                     /product="beta-tublulin"
     CDS             join(<153. .176,238. .360,427. .>537)
                     /codon_start=1
                     /product="beta-tublulin"
                     /protein_id="AAW57954.1"
                     /db_xref="GI:57901571"
                     /translation="VHLQTGQCGNQIGAAFWQTISGEHGLDSNGVYNGTSELQLERMS
                     VYFNEASGNKYVPRAVLVDLEPGTMDAVRAGPFGQLFRPDNF"
ORIGIN

Alignment Scores:
Pred. No.:                133           Length:         557
Score:                    42.00         Matches:        6
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              100.00%       Indels:         0
DB:                       15            Gaps:           0

US-09-858-332G-28 (1-6) x AY677225 (1-557)
```

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         47 CACAACCACAACCACAAC 64

RESULT 107
AY514428/c
LOCUS       AY514428                561 bp    DNA     linear   MAM 01-FEB-2004
DEFINITION  Bubalus bubalis alpha-S2 casein gene, exons 4 through 6.
ACCESSION   AY514428
VERSION     AY514428.1  GI:41350179
KEYWORDS    .
SOURCE      Bubalus bubalis (water buffalo)
  ORGANISM  Bubalus bubalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bubalus.
REFERENCE   1  (bases 1 to 561)
  AUTHORS   Misra,S.S., Sharma,A., Bhattacharya,T.K. and Kumar,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-2003) Animal Genetics Division, Indian Veterinary
            Research Institute, Bareilly, U.P. 243 122, India
FEATURES             Location/Qualifiers
     source          1. .561
                     /organism="Bubalus bubalis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:89462"
                     /sex="female"
                     /cell_type="blood"
                     /note="breed: Bhadawari"
     mRNA            join(<1. .24,109. .155,535. .>561)
                     /product="alpha-S2 casein"
                     /note="coding region not determined"
     exon            <1. .24
                     /number=4
     exon            109. .155
                     /number=5
     exon            535. .>561
                     /number=6
ORIGIN

Alignment Scores:
Pred. No.:                134           Length:         561
Score:                    42.00         Matches:        6
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              100.00%       Indels:         0
DB:                       4             Gaps:           0

US-09-858-332G-28 (1-6) x AY514428 (1-561)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        261 CACAATCACAATCATAAT 244

RESULT 108
BV412836/c
LOCUS       BV412836                569 bp    DNA     linear   STS 01-FEB-2005
DEFINITION  S229P6385RF4.T0 GermanShepherd Canis familiaris STS genomic,
            sequence tagged site.
ACCESSION   BV412836
VERSION     BV412836.1  GI:57786617
KEYWORDS    STS.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
            Canis.
REFERENCE   1  (bases 1 to 569)
  AUTHORS   Lindblad-Toh,K.
  TITLE     The genome sequence of Canis familiaris
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Kerstin Lindblad-Toh
            Whitehead Institute for Biomedical Research, Center for Genome
            Research
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172521477
            Fax: 6172580903
            Email: kersli@genome.wi.mit.edu
            Primer A: No sequence submitted
            Primer B: No sequence submitted
            STS size: 569
            Protocol:
            WGS-discovery (WGS):
            Paired-end low-coverage whole genome shotgun reads were generated
            from 9 breeds
            (German Shepherd, Rottweiler, Bedlington Terrier, Beagle, Labrador
            Retriever, English
            Shepherd, Italian Greyhound, Alaskan Malamute and the Portuguese
            Water Dog -100,000 each)
            and five other canids (Chinese, Alaskan, Indian and Spanish Gray
            Wolf as well as the
            Californian Coyote).
            The WGS reads were placed uniquely on the CanFam1.0 boxer assembly
            and SNP detection was
            carried out by SSAHA-SNP. 863872 reads were annotated as STSs and
            485941 SNPs were
            annotated with alleles from the boxer and the breed or canid from
            which the particular
            read came. The validation rate for these SNPs was estimated at
            approximately 98%.
            WGA-discovery (WGA) of Boxer/Poodle SNPs:
            A second set of SNPs was generated using a similar methodology
            except that the contigs
            from the 1.5x poodle assembly (Kirkness 2003) were used instead of
            WGS reads. Since this
            sequence lacked base quality scores, arbitrary quality scores of
            phred 40 were assigned
            before the poddle seqeunce was placed uniquely on the CanFam1.0
            boxer assembly and SNP
            detection was carried out by SSAHA-SNP. 1637780 SNPs were annotated
            with alleles from the
            boxer and the poodle. The validation rate for these SNPs was
            estimated at approximately TBD%.
            Internal-WGA-discovery (I-WGA):
            A third set of SNPs were discovered by comparing reads in the WGA
            assembly. SNPs were
            defined as mismatch positions that had a base quality of >= 30 on
            both reads in a region
            that aligned without gaps, and with at most one additional mismatch
            in the ten flanking
            bases. For each allele, at least one additional read had to confirm
            it. 731476 SNPs were
            annotated with alleles between the two boxer alleles. The
            validation rate for these SNPs
            was estimated at approximately TBD%.
FEATURES             Location/Qualifiers
     source          1. .569
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /strain="GermanShepherd"
                     /db_xref="taxon:9615"
                     /map="+ 2 22-515 81341124-81340632"
                     /clone_lib="GermanShepherd"
     STS             <1. .>569
ORIGIN

Alignment Scores:
Pred. No.:                136           Length:         569
Score:                    42.00         Matches:        6
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              100.00%       Indels:         0
DB:                       10            Gaps:           0
```

```
US-09-858-332G-28 (1-6) x BV412836 (1-569)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        563 CACAACCACAACCACAAC 546

RESULT 109
AX813675
LOCUS       AX813675                573 bp    DNA     linear   PAT 02-DEC-2003
DEFINITION  Sequence 8 from Patent WO03062415.
ACCESSION   AX813675
VERSION     AX813675.1  GI:38636043
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Drabek,D., Dekker,S. and Grosveld,F.G.
  TITLE     Transgenic animal and methods
  JOURNAL   Patent: WO 03062415-A 8 31-JUL-2003;
            Erasmus Unviersiteit (NL)
FEATURES             Location/Qualifiers
     source          1. .573
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="anti p15 A4"
ORIGIN

Alignment Scores:
Pred. No.:              137           Length:        573
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-28 (1-6) x AX813675 (1-573)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        473 CACAACCACAACCACAAC 490

RESULT 110
AX813678
LOCUS       AX813678                580 bp    DNA     linear   PAT 02-DEC-2003
DEFINITION  Sequence 11 from Patent WO03062415.
ACCESSION   AX813678
VERSION     AX813678.1  GI:38636046
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Drabek,D., Dekker,S. and Grosveld,F.G.
  TITLE     Transgenic animal and methods
  JOURNAL   Patent: WO 03062415-A 11 31-JUL-2003;
            Erasmus Unviersiteit (NL)
FEATURES             Location/Qualifiers
     source          1. .580
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="anti p15 C1"
ORIGIN

Alignment Scores:
Pred. No.:              138           Length:        580
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-28 (1-6) x AX813678 (1-580)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        480 CACAACCACAACCACAAC 497

RESULT 111
AX813680
LOCUS       AX813680                580 bp    DNA     linear   PAT 02-DEC-2003
DEFINITION  Sequence 13 from Patent WO03062415.
ACCESSION   AX813680
VERSION     AX813680.1  GI:38636048
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Drabek,D., Dekker,S. and Grosveld,F.G.
  TITLE     Transgenic animal and methods
  JOURNAL   Patent: WO 03062415-A 13 31-JUL-2003;
            Erasmus Unviersiteit (NL)
FEATURES             Location/Qualifiers
     source          1. .580
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="anti p15 H2"
ORIGIN

Alignment Scores:
Pred. No.:              138           Length:        580
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-28 (1-6) x AX813680 (1-580)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        480 CACAACCACAACCACAAC 497

RESULT 112
AR568827/c
LOCUS       AR568827                587 bp    DNA     linear   PAT 14-DEC-2004
DEFINITION  Sequence 1 from patent US 6740747.
ACCESSION   AR568827
VERSION     AR568827.1  GI:56568469
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 587)
  AUTHORS   Kaushik,A. and Saini,S.S.
  TITLE     Bovine VDJ cassette, BF1H1, suitable for antigenization
  JOURNAL   Patent: US 6740747-A 1 25-MAY-2004;
FEATURES             Location/Qualifiers
     source          1. .587
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              140           Length:        587
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-28 (1-6) x AR568827 (1-587)
```

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        411 CATAACCATAACCACAAC 394

RESULT 113
AX813681
LOCUS       AX813681                589 bp    DNA     linear   PAT 02-DEC-2003
DEFINITION  Sequence 14 from Patent WO03062415.
ACCESSION   AX813681
VERSION     AX813681.1  GI:38636049
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Drabek,D., Dekker,S. and Grosveld,F.G.
  TITLE     Transgenic animal and methods
  JOURNAL   Patent: WO 03062415-A 14 31-JUL-2003;
            Erasmus Unviersiteit (NL)
FEATURES             Location/Qualifiers
     source          1. .589
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Anti p27 D2"
ORIGIN

Alignment Scores:
Pred. No.:              141           Length:         589
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     6             Gaps:           0

US-09-858-332G-28 (1-6) x AX813681 (1-589)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        489 CACAACCACAACCACAAC 506

RESULT 114
BV276443
LOCUS       BV276443                591 bp    DNA     linear   STS 22-JAN-2005
DEFINITION  S232P6167RA4.T0 Beagle Canis familiaris STS genomic, sequence
            tagged site.
ACCESSION   BV276443
VERSION     BV276443.1  GI:57438672
KEYWORDS    STS.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
            Canis.
REFERENCE   1  (bases 1 to 591)
  AUTHORS   Lindblad-Toh,K.
  TITLE     The genome sequence of Canis familiaris
  JOURNAL   Unpublished (2004)
COMMENT
            Contact: Kerstin Lindblad-Toh
            Whitehead Institute for Biomedical Research, Center for Genome
            Research
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172521477
            Fax: 6172580903
            Email: kersli@genome.wi.mit.edu
            Primer A: No sequence submitted
            Primer B: No sequence submitted
            STS size: 591
            Protocol:
            WGS-discovery (WGS):
            Paired-end low-coverage whole genome shotgun reads were generated
            from 9 breeds
            (German Shepherd, Rottweiler, Bedlington Terrier, Beagle, Labrador
            Retriever, English
            Shepherd, Italian Greyhound, Alaskan Malamute and the Portuguese
            Water Dog -100,000 each)
            and five other canids (Chinese, Alaskan, Indian and Spanish Gray
            Wolf as well as the
            Californian Coyote).
            The WGS reads were placed uniquely on the CanFam1.0 boxer assembly
            and SNP detection was
            carried out by SSAHA-SNP. 863872 reads were annotated as STSs and
            485941 SNPs were
            annotated with alleles from the boxer and the breed or canid from
            which the particular
            read came. The validation rate for these SNPs was estimated at
            approximately 98%.
            WGA-discovery (WGA) of Boxer/Poodle SNPs:
            A second set of SNPs was generated using a similar methodology
            except that the contigs
            from the 1.5x poodle assembly (Kirkness 2003) were used instead of
            WGS reads. Since this
            sequence lacked base quality scores, arbitrary quality scores of
            phred 40 were assigned
            before the poddle seqeunce was placed uniquely on the CanFam1.0
            boxer assembly and SNP
            detection was carried out by SSAHA-SNP. 1637780 SNPs were annotated
            with alleles from the
            boxer and the poodle. The validation rate for these SNPs was
            estimated at approximately TBD%.
            Internal-WGA-discovery (I-WGA):
            A third set of SNPs were discovered by comparing reads in the WGA
            assembly. SNPs were
            defined as mismatch positions that had a base quality of >= 30 on
            both reads in a region
            that aligned without gaps, and with at most one additional mismatch
            in the ten flanking
            bases. For each allele, at least one additional read had to confirm
            it. 731476 SNPs were
            annotated with alleles between the two boxer alleles. The
            validation rate for these SNPs
            was estimated at approximately TBD%.
FEATURES             Location/Qualifiers
     source          1. .591
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /strain="Beagle"
                     /db_xref="taxon:9615"
                     /map="+ 18 8-531 52703068-52703589"
                     /clone_lib="Beagle"
     STS             <1. .>591
ORIGIN

Alignment Scores:
Pred. No.:              141           Length:         591
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     10            Gaps:           0

US-09-858-332G-28 (1-6) x BV276443 (1-591)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        523 CATAATCATAATCATAAT 540

RESULT 115
AX813682
LOCUS       AX813682                592 bp    DNA     linear   PAT 02-DEC-2003
DEFINITION  Sequence 15 from Patent WO03062415.
ACCESSION   AX813682
VERSION     AX813682.1  GI:38636050
KEYWORDS    .
```

```
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Drabek,D., Dekker,S. and Grosveld,F.G.
  TITLE     Transgenic animal and methods
  JOURNAL   Patent: WO 03062415-A 15 31-JUL-2003;
            Erasmus Unviersiteit (NL)
FEATURES             Location/Qualifiers
     source          1. .592
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="anti p27 G12"
ORIGIN

Alignment Scores:
Pred. No.:              141          Length:        592
Score:                  42.00        Matches:       6
Percent Similarity:     100.00%      Conservative:  0
Best Local Similarity:  100.00%      Mismatches:    0
Query Match:            100.00%      Indels:        0
DB:                     6            Gaps:          0

US-09-858-332G-28 (1-6) x AX813682 (1-592)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        492 CACAACCACAACCACAAC 509

RESULT 116
AR546074
LOCUS       AR546074                600 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 1205 from patent US 6747137.
ACCESSION   AR546074
VERSION     AR546074.1  GI:53939249
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 1205 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .600
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              143          Length:        600
Score:                  42.00        Matches:       6
Percent Similarity:     100.00%      Conservative:  0
Best Local Similarity:  100.00%      Mismatches:    0
Query Match:            100.00%      Indels:        0
DB:                     6            Gaps:          0

US-09-858-332G-28 (1-6) x AR546074 (1-600)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        194 CACAATCACAATCACAAT 211

RESULT 117
AR664528
LOCUS       AR664528                601 bp    DNA     linear   PAT 13-JUN-2005
DEFINITION  Sequence 5569 from patent US 6900016.
ACCESSION   AR664528
VERSION     AR664528.1  GI:67601308
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 601)
  AUTHORS   Venter,J.C., Zhang,J.N., Liu,X., Rowe,W., Cravchik,A., Kalush,F.,
            Naik,A., Subramanian,G. and Woodage,T.
  TITLE     Polymorphisms in known genes associated with inflammatory
            autoimmune disease, methods of detection and uses thereof
  JOURNAL   Patent: US 6900016-A 5569 31-MAY-2005;
            Applera Corporation; Norwalk, CT
FEATURES             Location/Qualifiers
     source          1. .601
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              143          Length:        601
Score:                  42.00        Matches:       6
Percent Similarity:     100.00%      Conservative:  0
Best Local Similarity:  100.00%      Mismatches:    0
Query Match:            100.00%      Indels:        0
DB:                     6            Gaps:          0

US-09-858-332G-28 (1-6) x AR664528 (1-601)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        277 CATAATCATAATCATAAT 294

RESULT 118
AR664529
LOCUS       AR664529                601 bp    DNA     linear   PAT 13-JUN-2005
DEFINITION  Sequence 5570 from patent US 6900016.
ACCESSION   AR664529
VERSION     AR664529.1  GI:67601309
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 601)
  AUTHORS   Venter,J.C., Zhang,J.N., Liu,X., Rowe,W., Cravchik,A., Kalush,F.,
            Naik,A., Subramanian,G. and Woodage,T.
  TITLE     Polymorphisms in known genes associated with inflammatory
            autoimmune disease, methods of detection and uses thereof
  JOURNAL   Patent: US 6900016-A 5570 31-MAY-2005;
            Applera Corporation; Norwalk, CT
FEATURES             Location/Qualifiers
     source          1. .601
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              143          Length:        601
Score:                  42.00        Matches:       6
Percent Similarity:     100.00%      Conservative:  0
Best Local Similarity:  100.00%      Mismatches:    0
Query Match:            100.00%      Indels:        0
DB:                     6            Gaps:          0

US-09-858-332G-28 (1-6) x AR664529 (1-601)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        307 CATAATCATAATCATAAT 324

RESULT 119
BV377419
LOCUS       BV377419                607 bp    DNA     linear   STS 28-JAN-2005
DEFINITION  S245P6200FA10.T0 SpanishGrayWolf Canis lupus STS genomic, sequence
            tagged site.
ACCESSION   BV377419
```

```
VERSION     BV377419.1  GI:57631489
KEYWORDS    STS.
SOURCE      Canis lupus (gray wolf)
  ORGANISM  Canis lupus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
            Canis.
REFERENCE   1  (bases 1 to 607)
  AUTHORS   Lindblad-Toh,K.
  TITLE     The genome sequence of Canis familiaris
  JOURNAL   Unpublished (2004)
COMMENT
            Contact: Kerstin Lindblad-Toh
            Whitehead Institute for Biomedical Research, Center for Genome
            Research
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172521477
            Fax: 6172580903
            Email: kersli@genome.wi.mit.edu
            Primer A: No sequence submitted
            Primer B: No sequence submitted
            STS size: 607
            Protocol:
            WGS-discovery (WGS):
            Paired-end low-coverage whole genome shotgun reads were generated
            from 9 breeds
            (German Shepherd, Rottweiler, Bedlington Terrier, Beagle, Labrador
            Retriever, English
            Shepherd, Italian Greyhound, Alaskan Malamute and the Portuguese
            Water Dog -100,000 each)
            and five other canids (Chinese, Alaskan, Indian and Spanish Gray
            Wolf as well as the
            Californian Coyote).
            The WGS reads were placed uniquely on the CanFam1.0 boxer assembly
            and SNP detection was
            carried out by SSAHA-SNP. 863872 reads were annotated as STSs and
            485941 SNPs were
            annotated with alleles from the boxer and the breed or canid from
            which the particular
            read came. The validation rate for these SNPs was estimated at
            approximately 98%.
            WGA-discovery (WGA) of Boxer/Poodle SNPs:
            A second set of SNPs was generated using a similar methodology
            except that the contigs
            from the 1.5x poodle assembly (Kirkness 2003) were used instead of
            WGS reads. Since this
            sequence lacked base quality scores, arbitrary quality scores of
            phred 40 were assigned
            before the poddle seqeunce was placed uniquely on the CanFam1.0
            boxer assembly and SNP
            detection was carried out by SSAHA-SNP. 1637780 SNPs were annotated
            with alleles from the
            boxer and the poodle. The validation rate for these SNPs was
            estimated at approximately TBD%.
            Internal-WGA-discovery (I-WGA):
            A third set of SNPs were discovered by comparing reads in the WGA
            assembly. SNPs were
            defined as mismatch positions that had a base quality of >= 30 on
            both reads in a region
            that aligned without gaps, and with at most one additional mismatch
            in the ten flanking
            bases. For each allele, at least one additional read had to confirm
            it. 731476 SNPs were
            annotated with alleles between the two boxer alleles. The
            validation rate for these SNPs
            was estimated at approximately TBD%.
FEATURES              Location/Qualifiers
     source           1. .607
                      /organism="Canis lupus"
                      /mol_type="genomic DNA"
                      /strain="SpanishGrayWolf"
                      /db_xref="taxon:9612"
                      /map="+ 9 22-528 20461941-20462447"
                      /clone_lib="SpanishGrayWolf"
     STS              <1. .>607
ORIGIN

Alignment Scores:
Pred. No.:                145             Length:          607
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       10              Gaps:            0

US-09-858-332G-28 (1-6) x BV377419 (1-607)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        554 CATAACCATAACCATAAC 571

RESULT 120
AR549202
LOCUS       AR549202                612 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 4333 from patent US 6747137.
ACCESSION   AR549202
VERSION     AR549202.1  GI:53942377
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 612)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 4333 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES              Location/Qualifiers
     source           1. .612
                      /organism="unknown"
                      /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:                146             Length:          612
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       6               Gaps:            0

US-09-858-332G-28 (1-6) x AR549202 (1-612)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        305 CACAACCACAACCACAAC 322

RESULT 121
AX386212
LOCUS       AX386212                616 bp    DNA     linear   PAT 19-MAR-2002
DEFINITION  Sequence 1140 from Patent WO0214500.
ACCESSION   AX386212
VERSION     AX386212.1  GI:19579342
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Escobedo,J., Garcia,P.D., Sudduth-Klinger,J., Reinhard,C.,
            Randazzo,F., Lamson,G., Scott,E.M., Zhang,G., Kassam,A., Pot,D. and
            Labat,I.
  TITLE     Human genes and gene expression products
  JOURNAL   Patent: WO 0214500-A 1140 21-FEB-2002;
            CHIRON CORPORATION (US) ; Hyseq Inc. (US)
```

```
FEATURES             Location/Qualifiers
     source          1. .616
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:              147          Length:        616
Score:                  42.00        Matches:       6
Percent Similarity:     100.00%      Conservative:  0
Best Local Similarity:  100.00%      Mismatches:    0
Query Match:            100.00%      Indels:        0
DB:                     6            Gaps:          0

US-09-858-332G-28 (1-6) x AX386212 (1-616)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         84 CACAACCACAACCACAAC 101

RESULT 122
BV229282
LOCUS       BV229282                619 bp    DNA     linear   STS 19-JAN-2005
DEFINITION  S233P6299FC9.T0 LabradorRetriever Canis familiaris STS genomic,
            sequence tagged site.
ACCESSION   BV229282
VERSION     BV229282.1  GI:57291820
KEYWORDS    STS.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
            Canis.
REFERENCE   1  (bases 1 to 619)
  AUTHORS   Lindblad-Toh,K.
  TITLE     The genome sequence of Canis familiaris
  JOURNAL   Unpublished (2004)
COMMENT
            Contact: Kerstin Lindblad-Toh
            Whitehead Institute for Biomedical Research, Center for Genome
            Research
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172521477
            Fax: 6172580903
            Email: kersli@genome.wi.mit.edu
            Primer A: No sequence submitted
            Primer B: No sequence submitted
            STS size: 619
            Protocol:
            WGS-discovery (WGS):
            Paired-end low-coverage whole genome shotgun reads were generated
            from 9 breeds
            (German Shepherd, Rottweiler, Bedlington Terrier, Beagle, Labrador
            Retriever, English
            Shepherd, Italian Greyhound, Alaskan Malamute and the Portuguese
            Water Dog -100,000 each)
            and five other canids (Chinese, Alaskan, Indian and Spanish Gray
            Wolf as well as the
            Californian Coyote).
            The WGS reads were placed uniquely on the CanFam1.0 boxer assembly
            and SNP detection was
            carried out by SSAHA-SNP. 863872 reads were annotated as STSs and
            485941 SNPs were
            annotated with alleles from the boxer and the breed or canid from
            which the particular
            read came. The validation rate for these SNPs was estimated at
            approximately 98%.
            WGA-discovery (WGA) of Boxer/Poodle SNPs:
            A second set of SNPs was generated using a similar methodology
            except that the contigs
            from the 1.5x poodle assembly (Kirkness 2003) were used instead of
            WGS reads. Since this
            sequence lacked base quality scores, arbitrary quality scores of
            phred 40 were assigned
            before the poddle seqeunce was placed uniquely on the CanFam1.0
            boxer assembly and SNP
            detection was carried out by SSAHA-SNP. 1637780 SNPs were annotated
            with alleles from the
            boxer and the poodle. The validation rate for these SNPs was
            estimated at approximately TBD%.
            Internal-WGA-discovery (I-WGA):
            A third set of SNPs were discovered by comparing reads in the WGA
            assembly. SNPs were
            defined as mismatch positions that had a base quality of >= 30 on
            both reads in a region
            that aligned without gaps, and with at most one additional mismatch
            in the ten flanking
            bases. For each allele, at least one additional read had to confirm
            it. 731476 SNPs were
            annotated with alleles between the two boxer alleles. The
            validation rate for these SNPs
            was estimated at approximately TBD%.
FEATURES             Location/Qualifiers
     source          1. .619
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /strain="LabradorRetriever"
                     /db_xref="taxon:9615"
                     /map="+ 13 22-571 61665551-61666099"
                     /clone_lib="LabradorRetriever"
     STS             <1. .>619
ORIGIN

Alignment Scores:
Pred. No.:              147          Length:        619
Score:                  42.00        Matches:       6
Percent Similarity:     100.00%      Conservative:  0
Best Local Similarity:  100.00%      Mismatches:    0
Query Match:            100.00%      Indels:        0
DB:                     10           Gaps:          0

US-09-858-332G-28 (1-6) x BV229282 (1-619)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         74 CATAATCATAATCATAAT 91

RESULT 123
BV268136/c
LOCUS       BV268136                620 bp    DNA     linear   STS 20-JAN-2005
DEFINITION  S235P6208FG12.T0 ItalianGreyhound Canis familiaris STS genomic,
            sequence tagged site.
ACCESSION   BV268136
VERSION     BV268136.1  GI:57330678
KEYWORDS    STS.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
            Canis.
REFERENCE   1  (bases 1 to 620)
  AUTHORS   Lindblad-Toh,K.
  TITLE     The genome sequence of Canis familiaris
  JOURNAL   Unpublished (2004)
COMMENT
            Contact: Kerstin Lindblad-Toh
            Whitehead Institute for Biomedical Research, Center for Genome
            Research
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172521477
            Fax: 6172580903
            Email: kersli@genome.wi.mit.edu
            Primer A: No sequence submitted
            Primer B: No sequence submitted
            STS size: 620
```

```
            Protocol:
            WGS-discovery (WGS):
            Paired-end low-coverage whole genome shotgun reads were generated
            from 9 breeds
            (German Shepherd, Rottweiler, Bedlington Terrier, Beagle, Labrador
            Retriever, English
            Shepherd, Italian Greyhound, Alaskan Malamute and the Portuguese
            Water Dog -100,000 each)
            and five other canids (Chinese, Alaskan, Indian and Spanish Gray
            Wolf as well as the
            Californian Coyote).
            The WGS reads were placed uniquely on the CanFam1.0 boxer assembly
            and SNP detection was
            carried out by SSAHA-SNP. 863872 reads were annotated as STSs and
            485941 SNPs were
            annotated with alleles from the boxer and the breed or canid from
            which the particular
            read came. The validation rate for these SNPs was estimated at
            approximately 98%.
            WGA-discovery (WGA) of Boxer/Poodle SNPs:
            A second set of SNPs was generated using a similar methodology
            except that the contigs
            from the 1.5x poodle assembly (Kirkness 2003) were used instead of
            WGS reads. Since this
            sequence lacked base quality scores, arbitrary quality scores of
            phred 40 were assigned
            before the poddle seqeunce was placed uniquely on the CanFam1.0
            boxer assembly and SNP
            detection was carried out by SSAHA-SNP. 1637780 SNPs were annotated
            with alleles from the
            boxer and the poodle. The validation rate for these SNPs was
            estimated at approximately TBD%.
            Internal-WGA-discovery (I-WGA):
            A third set of SNPs were discovered by comparing reads in the WGA
            assembly. SNPs were
            defined as mismatch positions that had a base quality of >= 30 on
            both reads in a region
            that aligned without gaps, and with at most one additional mismatch
            in the ten flanking
            bases. For each allele, at least one additional read had to confirm
            it. 731476 SNPs were
            annotated with alleles between the two boxer alleles. The
            validation rate for these SNPs
            was estimated at approximately TBD%.
FEATURES             Location/Qualifiers
     source          1. .620
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /strain="ItalianGreyhound"
                     /db_xref="taxon:9615"
                     /map="+ 1 22-571 25259576-25259027"
                     /clone_lib="ItalianGreyhound"
     STS             <1. .>620
ORIGIN

Alignment Scores:
Pred. No.:                148           Length:          620
Score:                    42.00         Matches:         6
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       10            Gaps:            0

US-09-858-332G-28 (1-6) x BV268136 (1-620)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         42 CATAATCATAATCATAAT 25

RESULT 124
BV260649
LOCUS       BV260649                 621 bp    DNA     linear   STS 22-JAN-2005
DEFINITION  S235P626RC8.T0 ItalianGreyhound Canis familiaris STS genomic,
            sequence tagged site.
ACCESSION   BV260649
VERSION     BV260649.1  GI:57323191
KEYWORDS    STS.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
            Canis.
REFERENCE   1  (bases 1 to 621)
  AUTHORS   Lindblad-Toh,K.
  TITLE     The genome sequence of Canis familiaris
  JOURNAL   Unpublished (2004)
COMMENT
            Contact: Kerstin Lindblad-Toh
            Whitehead Institute for Biomedical Research, Center for Genome
            Research
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172521477
            Fax: 6172580903
            Email: kersli@genome.wi.mit.edu
            Primer A: No sequence submitted
            Primer B: No sequence submitted
            STS size: 621
            Protocol:
            WGS-discovery (WGS):
            Paired-end low-coverage whole genome shotgun reads were generated
            from 9 breeds
            (German Shepherd, Rottweiler, Bedlington Terrier, Beagle, Labrador
            Retriever, English
            Shepherd, Italian Greyhound, Alaskan Malamute and the Portuguese
            Water Dog -100,000 each)
            and five other canids (Chinese, Alaskan, Indian and Spanish Gray
            Wolf as well as the
            Californian Coyote).
            The WGS reads were placed uniquely on the CanFam1.0 boxer assembly
            and SNP detection was
            carried out by SSAHA-SNP. 863872 reads were annotated as STSs and
            485941 SNPs were
            annotated with alleles from the boxer and the breed or canid from
            which the particular
            read came. The validation rate for these SNPs was estimated at
            approximately 98%.
            WGA-discovery (WGA) of Boxer/Poodle SNPs:
            A second set of SNPs was generated using a similar methodology
            except that the contigs
            from the 1.5x poodle assembly (Kirkness 2003) were used instead of
            WGS reads. Since this
            sequence lacked base quality scores, arbitrary quality scores of
            phred 40 were assigned
            before the poddle seqeunce was placed uniquely on the CanFam1.0
            boxer assembly and SNP
            detection was carried out by SSAHA-SNP. 1637780 SNPs were annotated
            with alleles from the
            boxer and the poodle. The validation rate for these SNPs was
            estimated at approximately TBD%.
            Internal-WGA-discovery (I-WGA):
            A third set of SNPs were discovered by comparing reads in the WGA
            assembly. SNPs were
            defined as mismatch positions that had a base quality of >= 30 on
            both reads in a region
            that aligned without gaps, and with at most one additional mismatch
            in the ten flanking
            bases. For each allele, at least one additional read had to confirm
            it. 731476 SNPs were
            annotated with alleles between the two boxer alleles. The
            validation rate for these SNPs
            was estimated at approximately TBD%.
FEATURES             Location/Qualifiers
     source          1. .621
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /strain="ItalianGreyhound"
```

```
                     /db_xref="taxon:9615"
                     /map="+ 23 17-570 7340735-7341288"
                     /clone_lib="ItalianGreyhound"
     STS             <1. .>621
ORIGIN

Alignment Scores:
Pred. No.:              148             Length:         621
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     10              Gaps:           0

US-09-858-332G-28 (1-6) x BV260649 (1-621)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        467 CATAACCATAACCATAAC 484

RESULT 125
BV542661
LOCUS       BV542661                 623 bp    DNA     linear   STS 09-APR-2005
DEFINITION  G591P57726RC1.T0 Clint Pan troglodytes verus STS genomic, sequence
            tagged site.
ACCESSION   BV542661
VERSION     BV542661.1  GI:62433681
KEYWORDS    STS.
SOURCE      Pan troglodytes verus
  ORGANISM  Pan troglodytes verus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1  (bases 1 to 623)
  AUTHORS   Mikkelsen,T.S., Hillier,W.L., Eichler,E.E., Zody,M.C. and
            Jaffe,D.B.
  TITLE     Initial Sequence of the Chimpanzee Genome and Comparison with the
            Human Genome
  JOURNAL   Unpublished (2005)
COMMENT
            Contact: Michael C. Zody
            Broad Institute of MIT and Harvard
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172580933
            Fax: 6172580903
            Email: mczody@broad.mit.edu
            Primer A: No sequence submitted
            Primer B: No sequence submitted
            STS size: 623
            Protocol:
            23,021,928 chimpanzee whole genome shotgun reads were aligned to
            the Human genome NCBI
            Build 34 (hg16,July 2003). Chimp WGS reads were from 9 donors,
            including Clint (Pan
            troglodytes verus), 3 other Pan troglodytes verus chimps
            (Donald,Karlien,Yvonne), 3 Pan
            troglodytes troglodytes chimps (Noemie,Masuku,Clara) and 2 chimps
            of unknown origin
            (Gon,Unknown Chimp). Common names:  Pan troglodytes verus is the
            western chimp and Pan
            troglodytes troglodytes is the central chimp. To be included in
            chimpanzee SNP discovery, a
            read must be at least 500bp in length, at least 50% of its base
            calls must have Phred
            score >= 20, at least 30% of its base calls must satisfy
            SNQS(30,25)(single strand NQS, the
            base in question has Phred score >= 30, the surrounding 10 bases in
            the read have Phred
            score >= 25), and the read must have at least 200 bp SNQS(30,25)
            bases. Reads not uniquely
            placed in the genome and read pairs whose two ends were not
            consistently placed were
            discarded. After above filtering, NQS(30,25) standard was applied
            to all pairs of
            overlapping reads to call NQS bases and SNPs. Alignments (between
            two reads) with less
            than 100 NQS bases or with SNP rate > 0.01 were discarded. To
            exclude alignment between two
            copies of a single read, comparisons between two reads that share
            95% of their genome
            alignments (>=95% bases of read A and >=95% bases of read B were
            placed at the same locus
            of human genome) were discarded.
FEATURES             Location/Qualifiers
     source          1. .623
                     /organism="Pan troglodytes verus"
                     /mol_type="genomic DNA"
                     /sub_species="verus"
                     /db_xref="taxon:37012"
                     /clone_lib="Clint"
     STS             <1. .>623
ORIGIN

Alignment Scores:
Pred. No.:              148             Length:         623
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     10              Gaps:           0

US-09-858-332G-28 (1-6) x BV542661 (1-623)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        272 CATAATCATAATCATAAT 289

RESULT 126
BV277306/c
LOCUS       BV277306                 628 bp    DNA     linear   STS 22-JAN-2005
DEFINITION  S232P6158RF4.T0 Beagle Canis familiaris STS genomic, sequence
            tagged site.
ACCESSION   BV277306
VERSION     BV277306.1  GI:57439535
KEYWORDS    STS.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
            Canis.
REFERENCE   1  (bases 1 to 628)
  AUTHORS   Lindblad-Toh,K.
  TITLE     The genome sequence of Canis familiaris
  JOURNAL   Unpublished (2004)
COMMENT
            Contact: Kerstin Lindblad-Toh
            Whitehead Institute for Biomedical Research, Center for Genome
            Research
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172521477
            Fax: 6172580903
            Email: kersli@genome.wi.mit.edu
            Primer A: No sequence submitted
            Primer B: No sequence submitted
            STS size: 628
            Protocol:
            WGS-discovery (WGS):
            Paired-end low-coverage whole genome shotgun reads were generated
            from 9 breeds
            (German Shepherd, Rottweiler, Bedlington Terrier, Beagle, Labrador
            Retriever, English
            Shepherd, Italian Greyhound, Alaskan Malamute and the Portuguese
            Water Dog -100,000 each)
            and five other canids (Chinese, Alaskan, Indian and Spanish Gray
            Wolf as well as the
            Californian Coyote).
```

```
            The WGS reads were placed uniquely on the CanFam1.0 boxer assembly
            and SNP detection was
            carried out by SSAHA-SNP. 863872 reads were annotated as STSs and
            485941 SNPs were
            annotated with alleles from the boxer and the breed or canid from
            which the particular
            read came. The validation rate for these SNPs was estimated at
            approximately 98%.
            WGA-discovery (WGA) of Boxer/Poodle SNPs:
            A second set of SNPs was generated using a similar methodology
            except that the contigs
            from the 1.5x poodle assembly (Kirkness 2003) were used instead of
            WGS reads. Since this
            sequence lacked base quality scores, arbitrary quality scores of
            phred 40 were assigned
            before the poddle seqeunce was placed uniquely on the CanFam1.0
            boxer assembly and SNP
            detection was carried out by SSAHA-SNP. 1637780 SNPs were annotated
            with alleles from the
            boxer and the poodle. The validation rate for these SNPs was
            estimated at approximately TBD%.
            Internal-WGA-discovery (I-WGA):
            A third set of SNPs were discovered by comparing reads in the WGA
            assembly. SNPs were
            defined as mismatch positions that had a base quality of >= 30 on
            both reads in a region
            that aligned without gaps, and with at most one additional mismatch
            in the ten flanking
            bases. For each allele, at least one additional read had to confirm
            it. 731476 SNPs were
            annotated with alleles between the two boxer alleles. The
            validation rate for these SNPs
            was estimated at approximately TBD%.
FEATURES             Location/Qualifiers
     source          1. .628
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /strain="Beagle"
                     /db_xref="taxon:9615"
                     /map="+ 22 22-587 58745776-58745211"
                     /clone_lib="Beagle"
     STS             <1. .>628
ORIGIN

Alignment Scores:
Pred. No.:                149             Length:         628
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       10              Gaps:           0

US-09-858-332G-28 (1-6) x BV277306 (1-628)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        246 CATAATCATAATCATAAT 229

RESULT 127
BV305608
LOCUS       BV305608                 632 bp    DNA     linear   STS 26-JAN-2005
DEFINITION  S236P6321RH5.T0 AlaskanMalamute Canis familiaris STS genomic,
            sequence tagged site.
ACCESSION   BV305608
VERSION     BV305608.1  GI:57467843
KEYWORDS    STS.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
            Canis.
REFERENCE   1  (bases 1 to 632)
  AUTHORS   Lindblad-Toh,K.
  TITLE     The genome sequence of Canis familiaris
  JOURNAL   Unpublished (2004)
COMMENT
            Contact: Kerstin Lindblad-Toh
            Whitehead Institute for Biomedical Research, Center for Genome
            Research
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172521477
            Fax: 6172580903
            Email: kersli@genome.wi.mit.edu
            Primer A: No sequence submitted
            Primer B: No sequence submitted
            STS size: 632
            Protocol:
            WGS-discovery (WGS):
            Paired-end low-coverage whole genome shotgun reads were generated
            from 9 breeds
            (German Shepherd, Rottweiler, Bedlington Terrier, Beagle, Labrador
            Retriever, English
            Shepherd, Italian Greyhound, Alaskan Malamute and the Portuguese
            Water Dog -100,000 each)
            and five other canids (Chinese, Alaskan, Indian and Spanish Gray
            Wolf as well as the
            Californian Coyote).
            The WGS reads were placed uniquely on the CanFam1.0 boxer assembly
            and SNP detection was
            carried out by SSAHA-SNP. 863872 reads were annotated as STSs and
            485941 SNPs were
            annotated with alleles from the boxer and the breed or canid from
            which the particular
            read came. The validation rate for these SNPs was estimated at
            approximately 98%.
            WGA-discovery (WGA) of Boxer/Poodle SNPs:
            A second set of SNPs was generated using a similar methodology
            except that the contigs
            from the 1.5x poodle assembly (Kirkness 2003) were used instead of
            WGS reads. Since this
            sequence lacked base quality scores, arbitrary quality scores of
            phred 40 were assigned
            before the poddle seqeunce was placed uniquely on the CanFam1.0
            boxer assembly and SNP
            detection was carried out by SSAHA-SNP. 1637780 SNPs were annotated
            with alleles from the
            boxer and the poodle. The validation rate for these SNPs was
            estimated at approximately TBD%.
            Internal-WGA-discovery (I-WGA):
            A third set of SNPs were discovered by comparing reads in the WGA
            assembly. SNPs were
            defined as mismatch positions that had a base quality of >= 30 on
            both reads in a region
            that aligned without gaps, and with at most one additional mismatch
            in the ten flanking
            bases. For each allele, at least one additional read had to confirm
            it. 731476 SNPs were
            annotated with alleles between the two boxer alleles. The
            validation rate for these SNPs
            was estimated at approximately TBD%.
FEATURES             Location/Qualifiers
     source          1. .632
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /strain="AlaskanMalamute"
                     /db_xref="taxon:9615"
                     /map="+ 3 33-570 28862467-28861930"
                     /clone_lib="AlaskanMalamute"
     STS             <1. .>632
ORIGIN

Alignment Scores:
Pred. No.:                150             Length:         632
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
```

```
Query Match:            100.00%        Indels:        0
DB:                     10             Gaps:          0

US-09-858-332G-28 (1-6) x BV305608 (1-632)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         63 CATAATCATAATCATAAT 80

RESULT 128
BV061248/c
LOCUS       BV061248                641 bp    DNA     linear   STS 31-MAY-2003
DEFINITION  S212P6042FC8.T0 CZECHII/Ei Mus musculus STS genomic, sequence
            tagged site.
ACCESSION   BV061248
VERSION     BV061248.1  GI:31177043
KEYWORDS    STS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 641)
  AUTHORS   Wade,C.M., Kulbokas,E.J. III, Kirby,A.W., Zody,M.C., Mullikin,J.C.,
            Lander,E.S., Lindblad-Toh,K. and Daly,M.J.
  TITLE     The mosaic structure of variation in the laboratory mouse genome
  JOURNAL   Nature 420 (6915), 574-578 (2002)
   PUBMED   12466852
COMMENT
            Contact: Kerstin Lindblad-Toh
            Whitehead Institute for Biomedical Research, Center for Genome
            Research
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172521477
            Fax: 6172580903
            Email: kersli@genome.wi.mit.edu
            Primer A: None
            Primer B: None
            STS size: 641
            Protocol:
            WGS-discovery: Paired-end low-coverage whole genome shotgun reads
            were generated from 129S1/SvImJ, C3H/HeJ, and BALB/cByJ. The WGS
            reads were placed uniquely on the MGSCv3 C57BL/6J assembly and SNP
            detection was carried out by SSAHA-SNP. 225,000 reads were
            annotated
            as STSs and 81,000 SNPs were annotated with alleles from C57BL/6J
            and the strain from which the particular read came. The validation
            rate for these SNPs was estimated at approximately 98%.
FEATURES              Location/Qualifiers
     source           1. .641
                      /organism="Mus musculus"
                      /mol_type="genomic DNA"
                      /strain="CZECHII/Ei"
                      /db_xref="taxon:10090"
                      /map="- 1 22-585 145186115-145185576"
                      /clone_lib="CZECHII/Ei"
     STS              <1. .>641
ORIGIN

Alignment Scores:
Pred. No.:              152            Length:        641
Score:                  42.00          Matches:       6
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     10             Gaps:          0

US-09-858-332G-28 (1-6) x BV061248 (1-641)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        508 CATAATCATAATCATAAT 491

RESULT 129
CNS019XM
LOCUS       CNS019XM                660 bp    mRNA    linear   PLN 20-MAR-2004
DEFINITION  Botrytis cinerea strain T4 cDNA library.
ACCESSION   AL112226
VERSION     AL112226.1  GI:5826845
KEYWORDS    cDNA library; EST, nitrate deprivation.
SOURCE      Botryotinia fuckeliana
  ORGANISM  Botryotinia fuckeliana
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomycetes;
            Helotiales; Sclerotiniaceae; Botryotinia.
REFERENCE   1  (bases 1 to 660)
  AUTHORS   Bitton,F., Levis,C., Fortini,D., Pradier,J.M. and Brygoo,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-1999) Phytopathologie, INRA, route de St Cyr,
            78026 Versailles, France
REFERENCE   2  (bases 1 to 660)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-1999) Genoscope - Centre National de Sequencage :
            CP 5706 91057 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     The cDNA library was made by using Botrytis cinerea T4 strain grown
            under nitrate deprivation conditions. The cDNA were cloned
            directionally in the pBSII vector.
FEATURES              Location/Qualifiers
     source           1. .660
                      /organism="Botryotinia fuckeliana"
                      /mol_type="mRNA"
                      /strain="T4"
                      /db_xref="taxon:40559"
                      /note="Genoscope sequence ID : W21D121"
ORIGIN

Alignment Scores:
Pred. No.:              157            Length:        660
Score:                  42.00          Matches:       6
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     15             Gaps:          0

US-09-858-332G-28 (1-6) x CNS019XM (1-660)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         42 CACAATCACAACCACAAT 59

RESULT 130
CNS01C21/c
LOCUS       CNS01C21                668 bp    mRNA    linear   PLN 20-MAR-2004
DEFINITION  Botrytis cinerea strain T4 cDNA library.
ACCESSION   AL114977
VERSION     AL114977.1  GI:5829596
KEYWORDS    cDNA library; EST, nitrate deprivation.
SOURCE      Botryotinia fuckeliana
  ORGANISM  Botryotinia fuckeliana
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomycetes;
            Helotiales; Sclerotiniaceae; Botryotinia.
REFERENCE   1  (bases 1 to 668)
  AUTHORS   Bitton,F., Levis,C., Fortini,D., Pradier,J.M. and Brygoo,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-1999) Phytopathologie, INRA, route de St Cyr,
            78026 Versailles, France
REFERENCE   2  (bases 1 to 668)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-1999) Genoscope - Centre National de Sequencage :
            CP 5706 91057 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     The cDNA library was made by using Botrytis cinerea T4 strain grown
            under nitrate deprivation conditions. The cDNA were cloned
            directionally in the pBSII vector.
```

```
FEATURES             Location/Qualifiers
     source          1. .668
                     /organism="Botryotinia fuckeliana"
                     /mol_type="mRNA"
                     /strain="T4"
                     /db_xref="taxon:40559"
                     /note="Genoscope sequence ID : W51B121"
ORIGIN

Alignment Scores:
Pred. No.:                159           Length:         668
Score:                    42.00         Matches:        6
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              100.00%       Indels:         0
DB:                       15            Gaps:           0

US-09-858-332G-28 (1-6) x CNS01C21 (1-668)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         55 CACAACCACAATCATAAC 38

RESULT 131
BV527818
LOCUS       BV527818                681 bp    DNA     linear   STS 08-APR-2005
DEFINITION  G591P611466FE10.T0 Clint Pan troglodytes verus STS genomic,
            sequence tagged site.
ACCESSION   BV527818
VERSION     BV527818.1  GI:62405602
KEYWORDS    STS.
SOURCE      Pan troglodytes verus
  ORGANISM  Pan troglodytes verus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1  (bases 1 to 681)
  AUTHORS   Mikkelsen,T.S., Hillier,W.L., Eichler,E.E., Zody,M.C. and
            Jaffe,D.B.
  TITLE     Initial Sequence of the Chimpanzee Genome and Comparison with the
            Human Genome
  JOURNAL   Unpublished (2005)
COMMENT
            Contact: Michael C. Zody
            Broad Institute of MIT and Harvard
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172580933
            Fax: 6172580903
            Email: mczody@broad.mit.edu
            Primer A: No sequence submitted
            Primer B: No sequence submitted
            STS size: 681
            Protocol:
            23,021,928 chimpanzee whole genome shotgun reads were aligned to
            the Human genome NCBI
            Build 34 (hg16,July 2003). Chimp WGS reads were from 9 donors,
            including Clint (Pan
            troglodytes verus), 3 other Pan troglodytes verus chimps
            (Donald,Karlien,Yvonne), 3 Pan
            troglodytes troglodytes chimps (Noemie,Masuku,Clara) and 2 chimps
            of unknown origin
            (Gon,Unknown Chimp). Common names:  Pan troglodytes verus is the
            western chimp and Pan
            troglodytes troglodytes is the central chimp. To be included in
            chimpanzee SNP discovery, a
            read must be at least 500bp in length, at least 50% of its base
            calls must have Phred
            score >= 20, at least 30% of its base calls must satisfy
            SNQS(30,25)(single strand NQS, the
            base in question has Phred score >= 30, the surrounding 10 bases in
            the read have Phred
            score >= 25), and the read must have at least 200 bp SNQS(30,25)
            bases. Reads not uniquely
            placed in the genome and read pairs whose two ends were not
            consistently placed were
            discarded. After above filtering, NQS(30,25) standard was applied
            to all pairs of
            overlapping reads to call NQS bases and SNPs. Alignments (between
            two reads) with less
            than 100 NQS bases or with SNP rate > 0.01 were discarded. To
            exclude alignment between two
            copies of a single read, comparisons between two reads that share
            95% of their genome
            alignments (>=95% bases of read A and >=95% bases of read B were
            placed at the same locus
            of human genome) were discarded.
FEATURES             Location/Qualifiers
     source          1. .681
                     /organism="Pan troglodytes verus"
                     /mol_type="genomic DNA"
                     /sub_species="verus"
                     /db_xref="taxon:37012"
                     /clone_lib="Clint"
     STS             <1. .>681
ORIGIN

Alignment Scores:
Pred. No.:                162           Length:         681
Score:                    42.00         Matches:        6
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              100.00%       Indels:         0
DB:                       10            Gaps:           0

US-09-858-332G-28 (1-6) x BV527818 (1-681)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        286 CATAATCATAATCATAAT 303

RESULT 132
BV540308/c
LOCUS       BV540308                686 bp    DNA     linear   STS 08-APR-2005
DEFINITION  G591P641708FB11.T0 Clint Pan troglodytes verus STS genomic,
            sequence tagged site.
ACCESSION   BV540308
VERSION     BV540308.1  GI:62418159
KEYWORDS    STS.
SOURCE      Pan troglodytes verus
  ORGANISM  Pan troglodytes verus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1  (bases 1 to 686)
  AUTHORS   Mikkelsen,T.S., Hillier,W.L., Eichler,E.E., Zody,M.C. and
            Jaffe,D.B.
  TITLE     Initial Sequence of the Chimpanzee Genome and Comparison with the
            Human Genome
  JOURNAL   Unpublished (2005)
COMMENT
            Contact: Michael C. Zody
            Broad Institute of MIT and Harvard
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172580933
            Fax: 6172580903
            Email: mczody@broad.mit.edu
            Primer A: No sequence submitted
            Primer B: No sequence submitted
            STS size: 686
            Protocol:
            23,021,928 chimpanzee whole genome shotgun reads were aligned to
            the Human genome NCBI
            Build 34 (hg16,July 2003). Chimp WGS reads were from 9 donors,
            including Clint (Pan
            troglodytes verus), 3 other Pan troglodytes verus chimps
            (Donald,Karlien,Yvonne), 3 Pan
```

```
            troglodytes troglodytes chimps (Noemie,Masuku,Clara) and 2 chimps
            of unknown origin
            (Gon,Unknown Chimp). Common names:  Pan troglodytes verus is the
            western chimp and Pan
            troglodytes troglodytes is the central chimp. To be included in
            chimpanzee SNP discovery, a
            read must be at least 500bp in length, at least 50% of its base
            calls must have Phred
            score >= 20, at least 30% of its base calls must satisfy
            SNQS(30,25)(single strand NQS, the
            base in question has Phred score >= 30, the surrounding 10 bases in
            the read have Phred
            score >= 25), and the read must have at least 200 bp SNQS(30,25)
            bases. Reads not uniquely
            placed in the genome and read pairs whose two ends were not
            consistently placed were
            discarded. After above filtering, NQS(30,25) standard was applied
            to all pairs of
            overlapping reads to call NQS bases and SNPs. Alignments (between
            two reads) with less
            than 100 NQS bases or with SNP rate > 0.01 were discarded. To
            exclude alignment between two
            copies of a single read, comparisons between two reads that share
            95% of their genome
            alignments (>=95% bases of read A and >=95% bases of read B were
            placed at the same locus
            of human genome) were discarded.
FEATURES             Location/Qualifiers
     source          1. .686
                     /organism="Pan troglodytes verus"
                     /mol_type="genomic DNA"
                     /sub_species="verus"
                     /db_xref="taxon:37012"
                     /clone_lib="Clint"
     STS             <1. .>686
ORIGIN

Alignment Scores:
Pred. No.:              163             Length:         686
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     10              Gaps:           0

US-09-858-332G-28 (1-6) x BV540308 (1-686)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         23 CATAATCATAATCATAAT 6

RESULT 133
BV541444
LOCUS       BV541444                687 bp    DNA     linear   STS 08-APR-2005
DEFINITION  G591P55831RG8.T0 Clint Pan troglodytes verus STS genomic, sequence
            tagged site.
ACCESSION   BV541444
VERSION     BV541444.1  GI:62419295
KEYWORDS    STS.
SOURCE      Pan troglodytes verus
  ORGANISM  Pan troglodytes verus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1  (bases 1 to 687)
  AUTHORS   Mikkelsen,T.S., Hillier,W.L., Eichler,E.E., Zody,M.C. and
            Jaffe,D.B.
  TITLE     Initial Sequence of the Chimpanzee Genome and Comparison with the
            Human Genome
  JOURNAL   Unpublished (2005)
COMMENT
            Contact: Michael C. Zody
            Broad Institute of MIT and Harvard
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172580933
            Fax: 6172580903
            Email: mczody@broad.mit.edu
            Primer A: No sequence submitted
            Primer B: No sequence submitted
            STS size: 687
            Protocol:
            23,021,928 chimpanzee whole genome shotgun reads were aligned to
            the Human genome NCBI
            Build 34 (hg16,July 2003). Chimp WGS reads were from 9 donors,
            including Clint (Pan
            troglodytes verus), 3 other Pan troglodytes verus chimps
            (Donald,Karlien,Yvonne), 3 Pan
            troglodytes troglodytes chimps (Noemie,Masuku,Clara) and 2 chimps
            of unknown origin
            (Gon,Unknown Chimp). Common names:  Pan troglodytes verus is the
            western chimp and Pan
            troglodytes troglodytes is the central chimp. To be included in
            chimpanzee SNP discovery, a
            read must be at least 500bp in length, at least 50% of its base
            calls must have Phred
            score >= 20, at least 30% of its base calls must satisfy
            SNQS(30,25)(single strand NQS, the
            base in question has Phred score >= 30, the surrounding 10 bases in
            the read have Phred
            score >= 25), and the read must have at least 200 bp SNQS(30,25)
            bases. Reads not uniquely
            placed in the genome and read pairs whose two ends were not
            consistently placed were
            discarded. After above filtering, NQS(30,25) standard was applied
            to all pairs of
            overlapping reads to call NQS bases and SNPs. Alignments (between
            two reads) with less
            than 100 NQS bases or with SNP rate > 0.01 were discarded. To
            exclude alignment between two
            copies of a single read, comparisons between two reads that share
            95% of their genome
            alignments (>=95% bases of read A and >=95% bases of read B were
            placed at the same locus
            of human genome) were discarded.
FEATURES             Location/Qualifiers
     source          1. .687
                     /organism="Pan troglodytes verus"
                     /mol_type="genomic DNA"
                     /sub_species="verus"
                     /db_xref="taxon:37012"
                     /clone_lib="Clint"
     STS             <1. .>687
ORIGIN

Alignment Scores:
Pred. No.:              163             Length:         687
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     10              Gaps:           0

US-09-858-332G-28 (1-6) x BV541444 (1-687)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        379 CATAATCATAATCACAAT 396

RESULT 134
AF187066
LOCUS       AF187066                700 bp    mRNA    linear   ROD 11-JUN-2000
DEFINITION  Mus musculus p75NTR-associated cell death executor (Nade) mRNA,
            complete cds.
ACCESSION   AF187066
VERSION     AF187066.1  GI:8452897
KEYWORDS    .
```

```
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 700)
  AUTHORS   Mukai,J., Hachiya,T., Shoji-Hoshino,S., Kimura,M.T., Nadano,D.,
            Suvanto,P., Hanaoka,T., Li,Y., Irie,S., Greene,L.A. and Sato,T.A.
  TITLE     NADE, a p75NTR-associated cell death executor, is involved in
            signal transduction mediated by the common neurotrophin receptor
            p75NTR
  JOURNAL   J. Biol. Chem. 275 (23), 17566-17570 (2000)
   PUBMED   10764727
REFERENCE   2  (bases 1 to 700)
  AUTHORS   Mukai,J., Hachiya,T., Hoshino,S., Kimura,M., Nadano,D., Suvanto,P.,
            Hanaoka,T., Li,Y., Irie,S. and Sato,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-1999) Otolaryngology/Pathology, Columbia
            University, 630 West 168th St., P&S 11-451, New York, NY 10032, USA
FEATURES             Location/Qualifiers
     source          1. .700
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="BALB/c"
                     /db_xref="taxon:10090"
     gene            1. .700
                     /gene="Nade"
     CDS             177. .551
                     /gene="Nade"
                     /note="involved in the common neurotrophin receptor
                     p75NTR-mediated signal transduction; NADE"
                     /codon_start=1
                     /product="p75NTR-associated cell death executor"
                     /protein_id="AAF75131.1"
                     /db_xref="GI:8452898"
                     /translation="MANVHQENEEMEQPLQNGQEDRPVGGGEGHQPAANNNNNNHNHN
                     HNHHRRGQARRLAPNFRWAIPNRQMNDGLGGDGDDMEMFMEEMREIRRKLRELQLRNC
                     LRILMGELSNHHDHHDEFCLMP"
ORIGIN

Alignment Scores:
Pred. No.:                166             Length:         700
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       9               Gaps:           0

US-09-858-332G-28 (1-6) x AF187066 (1-700)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        297 CACAACCATAACCACAAC 314

RESULT 135
BV533882
LOCUS       BV533882                 703 bp    DNA     linear   STS 08-APR-2005
DEFINITION  G591P625286RG10.T0 Clint Pan troglodytes verus STS genomic,
            sequence tagged site.
ACCESSION   BV533882
VERSION     BV533882.1  GI:62411666
KEYWORDS    STS.
SOURCE      Pan troglodytes verus
  ORGANISM  Pan troglodytes verus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1  (bases 1 to 703)
  AUTHORS   Mikkelsen,T.S., Hillier,W.L., Eichler,E.E., Zody,M.C. and
            Jaffe,D.B.
  TITLE     Initial Sequence of the Chimpanzee Genome and Comparison with the
            Human Genome
  JOURNAL   Unpublished (2005)
COMMENT
            Contact: Michael C. Zody
            Broad Institute of MIT and Harvard
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172580933
            Fax: 6172580903
            Email: mczody@broad.mit.edu
            Primer A: No sequence submitted
            Primer B: No sequence submitted
            STS size: 703
            Protocol:
            23,021,928 chimpanzee whole genome shotgun reads were aligned to
            the Human genome NCBI
            Build 34 (hg16,July 2003). Chimp WGS reads were from 9 donors,
            including Clint (Pan
            troglodytes verus), 3 other Pan troglodytes verus chimps
            (Donald,Karlien,Yvonne), 3 Pan
            troglodytes troglodytes chimps (Noemie,Masuku,Clara) and 2 chimps
            of unknown origin
            (Gon,Unknown Chimp). Common names:  Pan troglodytes verus is the
            western chimp and Pan
            troglodytes troglodytes is the central chimp. To be included in
            chimpanzee SNP discovery, a
            read must be at least 500bp in length, at least 50% of its base
            calls must have Phred
            score >= 20, at least 30% of its base calls must satisfy
            SNQS(30,25)(single strand NQS, the
            base in question has Phred score >= 30, the surrounding 10 bases in
            the read have Phred
            score >= 25), and the read must have at least 200 bp SNQS(30,25)
            bases. Reads not uniquely
            placed in the genome and read pairs whose two ends were not
            consistently placed were
            discarded. After above filtering, NQS(30,25) standard was applied
            to all pairs of
            overlapping reads to call NQS bases and SNPs. Alignments (between
            two reads) with less
            than 100 NQS bases or with SNP rate > 0.01 were discarded. To
            exclude alignment between two
            copies of a single read, comparisons between two reads that share
            95% of their genome
            alignments (>=95% bases of read A and >=95% bases of read B were
            placed at the same locus
            of human genome) were discarded.
FEATURES             Location/Qualifiers
     source          1. .703
                     /organism="Pan troglodytes verus"
                     /mol_type="genomic DNA"
                     /sub_species="verus"
                     /db_xref="taxon:37012"
                     /clone_lib="Clint"
     STS             <1. .>703
ORIGIN

Alignment Scores:
Pred. No.:                167             Length:         703
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       10              Gaps:           0

US-09-858-332G-28 (1-6) x BV533882 (1-703)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        539 CATAATCATAATCATAAT 556

RESULT 136
AR549947
LOCUS       AR549947                 708 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 5078 from patent US 6747137.
ACCESSION   AR549947
```

```
VERSION     AR549947.1  GI:53943122
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 708)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 5078 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .708
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              168           Length:        708
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-28 (1-6) x AR549947 (1-708)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        425 CACAACCACAACCACAAC 442

RESULT 137
AY617399
LOCUS       AY617399                720 bp    DNA     linear   INV 20-OCT-2004
DEFINITION  Sterkiella histriomuscorum clone UGC1O0002_E06_R genomic sequence.
ACCESSION   AY617399
VERSION     AY617399.1  GI:47175819
KEYWORDS    .
SOURCE      Sterkiella histriomuscorum (Oxytricha trifallax)
  ORGANISM  Sterkiella histriomuscorum
            Eukaryota; Alveolata; Ciliophora; Spirotrichea; Stichotrichia;
            Stichotrichida; Oxytrichidae; Sterkiella.
REFERENCE   1  (bases 1 to 720)
  AUTHORS   Doak,T.G., Cavalcanti,A.R., Stover,N.A., Dunn,D.M., Weiss,R.,
            Herrick,G. and Landweber,L.F.
  TITLE     Sequencing the Oxytricha trifallax macronuclear genome: a pilot
            project
  JOURNAL   Trends Genet. 19 (11), 603-607 (2003)
   PUBMED   14585610
REFERENCE   2  (bases 1 to 720)
  AUTHORS   Cavalcanti,A.R.O., Dunn,D.M., Weiss,R., Herrick,G., Landweber,L.F.
            and Doak,T.G.
  TITLE     Sequence features of Oxytricha trifallax (class Spirotrichea)
            macronuclear telomeric and subtelomeric sequences
  JOURNAL   Protist 155 (3), 311-322 (2004)
REFERENCE   3  (bases 1 to 720)
  AUTHORS   Cavalcanti,A.R., Stover,N.A., Orecchia,L., Doak,T.G. and
            Landweber,L.F.
  TITLE     Coding properties of Oxytricha trifallax (Sterkiella
            histriomuscorum) macronuclear chromosomes: analysis of a pilot
            genome project
  JOURNAL   Chromosoma 113 (2), 69-76 (2004)
   PUBMED   15258807
REFERENCE   4  (bases 1 to 720)
  AUTHORS   Doak,T.G., Cavalcanti,A.R.O., Stover,N.A., Orecchia,L. and
            Landweber,L.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-APR-2004) EEB Department, Princeton University,
            Princeton, NJ 08544, USA
FEATURES             Location/Qualifiers
     source          1. .720
                     /organism="Sterkiella histriomuscorum"
                     /mol_type="genomic DNA"
                     /strain="SLC172"
                     /db_xref="taxon:94289"
                     /clone="UGC1O0002_E06_R"
                     /note="one-pass macronuclear genomic sequence; offspring
                     of a crossing between strains JRB310 and JRB510"
     repeat_region   1. .20
                     /note="telomeric_repeat"
                     /rpt_type=tandem
     misc_feature    94. .720
                     /note="similarity to gi|1168519|sp|P04931|ARP_PLAFA
                     ASPARAGINE-RICH PROTEIN (AG319) (ARP) with blast E value <
                     1.00E-5"
                     /evidence=not_experimental
ORIGIN

Alignment Scores:
Pred. No.:              171           Length:        720
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     2             Gaps:          0

US-09-858-332G-28 (1-6) x AY617399 (1-720)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        703 CACAATCATAATCATAAC 720

RESULT 138
AR546135
LOCUS       AR546135                726 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 1266 from patent US 6747137.
ACCESSION   AR546135
VERSION     AR546135.1  GI:53939310
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 726)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 1266 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .726
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              172           Length:        726
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-28 (1-6) x AR546135 (1-726)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         64 CATAACCATAATCACAAT 81

RESULT 139
BV502973/c
LOCUS       BV502973                737 bp    DNA     linear   STS 07-APR-2005
DEFINITION  qkb91c12.g1 Clint Pan troglodytes verus STS genomic, sequence
            tagged site.
ACCESSION   BV502973
VERSION     BV502973.1  GI:62346715
KEYWORDS    STS.
```

```
SOURCE      Pan troglodytes verus
  ORGANISM  Pan troglodytes verus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1  (bases 1 to 737)
  AUTHORS   Mikkelsen,T.S., Hillier,W.L., Eichler,E.E., Zody,M.C. and
            Jaffe,D.B.
  TITLE     Initial Sequence of the Chimpanzee Genome and Comparison with the
            Human Genome
  JOURNAL   Unpublished (2005)
COMMENT
            Contact: Michael C. Zody
            Broad Institute of MIT and Harvard
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172580933
            Fax: 6172580903
            Email: mczody@broad.mit.edu
            Primer A: No sequence submitted
            Primer B: No sequence submitted
            STS size: 737
            Protocol:
            23,021,928 chimpanzee whole genome shotgun reads were aligned to
            the Human genome NCBI
            Build 34 (hg16,July 2003). Chimp WGS reads were from 9 donors,
            including Clint (Pan
            troglodytes verus), 3 other Pan troglodytes verus chimps
            (Donald,Karlien,Yvonne), 3 Pan
            troglodytes troglodytes chimps (Noemie,Masuku,Clara) and 2 chimps
            of unknown origin
            (Gon,Unknown Chimp). Common names:  Pan troglodytes verus is the
            western chimp and Pan
            troglodytes troglodytes is the central chimp. To be included in
            chimpanzee SNP discovery, a
            read must be at least 500bp in length, at least 50% of its base
            calls must have Phred
            score >= 20, at least 30% of its base calls must satisfy
            SNQS(30,25)(single strand NQS, the
            base in question has Phred score >= 30, the surrounding 10 bases in
            the read have Phred
            score >= 25), and the read must have at least 200 bp SNQS(30,25)
            bases. Reads not uniquely
            placed in the genome and read pairs whose two ends were not
            consistently placed were
            discarded. After above filtering, NQS(30,25) standard was applied
            to all pairs of
            overlapping reads to call NQS bases and SNPs. Alignments (between
            two reads) with less
            than 100 NQS bases or with SNP rate > 0.01 were discarded. To
            exclude alignment between two
            copies of a single read, comparisons between two reads that share
            95% of their genome
            alignments (>=95% bases of read A and >=95% bases of read B were
            placed at the same locus
            of human genome) were discarded.
FEATURES              Location/Qualifiers
     source           1. .737
                      /organism="Pan troglodytes verus"
                      /mol_type="genomic DNA"
                      /sub_species="verus"
                      /db_xref="taxon:37012"
                      /clone_lib="Clint"
     STS              <1. .>737
ORIGIN

Alignment Scores:
Pred. No.:                174             Length:         737
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       10              Gaps:           0
```

```
US-09-858-332G-28 (1-6) x BV502973 (1-737)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        223 CATAATCATAATCACAAT 206

RESULT 140
BV553142
LOCUS       BV553142                 742 bp    DNA     linear   STS 09-APR-2005
DEFINITION  S217P6664FE4.T0 Noemie Pan troglodytes troglodytes STS genomic,
            sequence tagged site.
ACCESSION   BV553142
VERSION     BV553142.1  GI:62444163
KEYWORDS    STS.
SOURCE      Pan troglodytes troglodytes
  ORGANISM  Pan troglodytes troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1  (bases 1 to 742)
  AUTHORS   Mikkelsen,T.S., Hillier,W.L., Eichler,E.E., Zody,M.C. and
            Jaffe,D.B.
  TITLE     Initial Sequence of the Chimpanzee Genome and Comparison with the
            Human Genome
  JOURNAL   Unpublished (2005)
COMMENT
            Contact: Michael C. Zody
            Broad Institute of MIT and Harvard
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172580933
            Fax: 6172580903
            Email: mczody@broad.mit.edu
            Primer A: No sequence submitted
            Primer B: No sequence submitted
            STS size: 742
            Protocol:
            23,021,928 chimpanzee whole genome shotgun reads were aligned to
            the Human genome NCBI
            Build 34 (hg16,July 2003). Chimp WGS reads were from 9 donors,
            including Clint (Pan
            troglodytes verus), 3 other Pan troglodytes verus chimps
            (Donald,Karlien,Yvonne), 3 Pan
            troglodytes troglodytes chimps (Noemie,Masuku,Clara) and 2 chimps
            of unknown origin
            (Gon,Unknown Chimp). Common names:  Pan troglodytes verus is the
            western chimp and Pan
            troglodytes troglodytes is the central chimp. To be included in
            chimpanzee SNP discovery, a
            read must be at least 500bp in length, at least 50% of its base
            calls must have Phred
            score >= 20, at least 30% of its base calls must satisfy
            SNQS(30,25)(single strand NQS, the
            base in question has Phred score >= 30, the surrounding 10 bases in
            the read have Phred
            score >= 25), and the read must have at least 200 bp SNQS(30,25)
            bases. Reads not uniquely
            placed in the genome and read pairs whose two ends were not
            consistently placed were
            discarded. After above filtering, NQS(30,25) standard was applied
            to all pairs of
            overlapping reads to call NQS bases and SNPs. Alignments (between
            two reads) with less
            than 100 NQS bases or with SNP rate > 0.01 were discarded. To
            exclude alignment between two
            copies of a single read, comparisons between two reads that share
            95% of their genome
            alignments (>=95% bases of read A and >=95% bases of read B were
            placed at the same locus
            of human genome) were discarded.
FEATURES              Location/Qualifiers
     source           1. .742
                      /organism="Pan troglodytes troglodytes"
                      /mol_type="genomic DNA"
```

```
                     /sub_species="troglodytes"
                     /db_xref="taxon:37011"
                     /clone_lib="Noemie"
     STS             <1. .>742
ORIGIN

Alignment Scores:
Pred. No.:                176             Length:         742
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       10              Gaps:           0

US-09-858-332G-28 (1-6) x BV553142 (1-742)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        496 CATAATCATAATCATAAT 513

RESULT 141
AR546771
LOCUS       AR546771                  747 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 1902 from patent US 6747137.
ACCESSION   AR546771
VERSION     AR546771.1  GI:53939946
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 747)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 1902 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .747
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:                177             Length:         747
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x AR546771 (1-747)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         91 CATAATCACAACCACAAT 108

RESULT 142
AR548753
LOCUS       AR548753                  747 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 3884 from patent US 6747137.
ACCESSION   AR548753
VERSION     AR548753.1  GI:53941928
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 747)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 3884 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .747
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:                177             Length:         747
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x AR548753 (1-747)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         36 CATAACCATAACCATAAC 53

RESULT 143
AR552017
LOCUS       AR552017                  768 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 7148 from patent US 6747137.
ACCESSION   AR552017
VERSION     AR552017.1  GI:53945192
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 768)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 7148 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .768
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:                181             Length:         768
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x AR552017 (1-768)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        266 CACAACCACAACCACAAT 283

RESULT 144
PM2C9G/c
LOCUS       PM2C9G                    786 bp    DNA     linear   STS 29-MAY-2003
DEFINITION  Penicillium marneffei STS, clone pm2c9.g, sequence tagged site.
ACCESSION   AL684736
VERSION     AL684736.1  GI:19336195
KEYWORDS    STS.
SOURCE      Penicillium marneffei
  ORGANISM  Penicillium marneffei
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
REFERENCE   1
  AUTHORS   Yuen,K.Y., Pascal,G., Wong,S.S., Glaser,P., Woo,P.C., Kunst,F.,
            Cai,J.J., Cheung,E.Y., Medigue,C. and Danchin,A.
  TITLE     Exploring the Penicillium marneffei genome
  JOURNAL   Arch. Microbiol. 179 (5), 339-353 (2003)
   PUBMED   12640520
REFERENCE   2  (bases 1 to 786)
```

```
  AUTHORS   Danchin,A. and Pascal,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-2002) Danchin A., HKU-Pasteur Research Centre,
            Dexter HC Man Building 8, Sassoon Road, Pokfulam, Hong Kong
FEATURES             Location/Qualifiers
     source          1. .786
                     /organism="Penicillium marneffei"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:37727"
                     /clone="pm2c9.g"
ORIGIN

Alignment Scores:
Pred. No.:                186             Length:         786
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       10              Gaps:           0

US-09-858-332G-28 (1-6) x PM2C9G (1-786)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        442 CACAACCACAACCACAAC 425

RESULT 145
AR556540/c
LOCUS       AR556540                 789 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 11671 from patent US 6747137.
ACCESSION   AR556540
VERSION     AR556540.1  GI:53949715
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 789)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 11671 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .789
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:                186             Length:         789
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x AR556540 (1-789)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        304 CATAATCATAATCATAAT 287

RESULT 146
BV509645/c
LOCUS       BV509645                 795 bp    DNA     linear   STS 07-APR-2005
DEFINITION  qwf38d09.g1 Clint Pan troglodytes verus STS genomic, sequence
            tagged site.
ACCESSION   BV509645
VERSION     BV509645.1  GI:62353387
KEYWORDS    STS.
SOURCE      Pan troglodytes verus
  ORGANISM  Pan troglodytes verus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1  (bases 1 to 795)
  AUTHORS   Mikkelsen,T.S., Hillier,W.L., Eichler,E.E., Zody,M.C. and
            Jaffe,D.B.
  TITLE     Initial Sequence of the Chimpanzee Genome and Comparison with the
            Human Genome
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Michael C. Zody
            Broad Institute of MIT and Harvard
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172580933
            Fax: 6172580903
            Email: mczody@broad.mit.edu
            Primer A: No sequence submitted
            Primer B: No sequence submitted
            STS size: 795
            Protocol:
            23,021,928 chimpanzee whole genome shotgun reads were aligned to
            the Human genome NCBI
            Build 34 (hg16,July 2003). Chimp WGS reads were from 9 donors,
            including Clint (Pan
            troglodytes verus), 3 other Pan troglodytes verus chimps
            (Donald,Karlien,Yvonne), 3 Pan
            troglodytes troglodytes chimps (Noemie,Masuku,Clara) and 2 chimps
            of unknown origin
            (Gon,Unknown Chimp). Common names:  Pan troglodytes verus is the
            western chimp and Pan
            troglodytes troglodytes is the central chimp. To be included in
            chimpanzee SNP discovery, a
            read must be at least 500bp in length, at least 50% of its base
            calls must have Phred
            score >= 20, at least 30% of its base calls must satisfy
            SNQS(30,25)(single strand NQS, the
            base in question has Phred score >= 30, the surrounding 10 bases in
            the read have Phred
            score >= 25), and the read must have at least 200 bp SNQS(30,25)
            bases. Reads not uniquely
            placed in the genome and read pairs whose two ends were not
            consistently placed were
            discarded. After above filtering, NQS(30,25) standard was applied
            to all pairs of
            overlapping reads to call NQS bases and SNPs. Alignments (between
            two reads) with less
            than 100 NQS bases or with SNP rate > 0.01 were discarded. To
            exclude alignment between two
            copies of a single read, comparisons between two reads that share
            95% of their genome
            alignments (>=95% bases of read A and >=95% bases of read B were
            placed at the same locus
            of human genome) were discarded.
FEATURES             Location/Qualifiers
     source          1. .795
                     /organism="Pan troglodytes verus"
                     /mol_type="genomic DNA"
                     /sub_species="verus"
                     /db_xref="taxon:37012"
                     /clone_lib="Clint"
     STS             <1. .>795
ORIGIN

Alignment Scores:
Pred. No.:                188             Length:         795
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       10              Gaps:           0

US-09-858-332G-28 (1-6) x BV509645 (1-795)

Qy          1 HisAsnHisAsnHisAsn 6
```

```
                ||||||||||||||||||
Db        202 CATAATCATAATCACAAT 185

RESULT 147
BV527033/c
LOCUS       BV527033                 810 bp    DNA     linear   STS 08-APR-2005
DEFINITION  G591P609151FA6.T0 Clint Pan troglodytes verus STS genomic, sequence
            tagged site.
ACCESSION   BV527033
VERSION     BV527033.1  GI:62404817
KEYWORDS    STS.
SOURCE      Pan troglodytes verus
  ORGANISM  Pan troglodytes verus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1  (bases 1 to 810)
  AUTHORS   Mikkelsen,T.S., Hillier,W.L., Eichler,E.E., Zody,M.C. and
            Jaffe,D.B.
  TITLE     Initial Sequence of the Chimpanzee Genome and Comparison with the
            Human Genome
  JOURNAL   Unpublished (2005)
COMMENT
            Contact: Michael C. Zody
            Broad Institute of MIT and Harvard
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172580933
            Fax: 6172580903
            Email: mczody@broad.mit.edu
            Primer A: No sequence submitted
            Primer B: No sequence submitted
            STS size: 810
            Protocol:
            23,021,928 chimpanzee whole genome shotgun reads were aligned to
            the Human genome NCBI
            Build 34 (hg16,July 2003). Chimp WGS reads were from 9 donors,
            including Clint (Pan
            troglodytes verus), 3 other Pan troglodytes verus chimps
            (Donald,Karlien,Yvonne), 3 Pan
            troglodytes troglodytes chimps (Noemie,Masuku,Clara) and 2 chimps
            of unknown origin
            (Gon,Unknown Chimp). Common names:  Pan troglodytes verus is the
            western chimp and Pan
            troglodytes troglodytes is the central chimp. To be included in
            chimpanzee SNP discovery, a
            read must be at least 500bp in length, at least 50% of its base
            calls must have Phred
            score >= 20, at least 30% of its base calls must satisfy
            SNQS(30,25)(single strand NQS, the
            base in question has Phred score >= 30, the surrounding 10 bases in
            the read have Phred
            score >= 25), and the read must have at least 200 bp SNQS(30,25)
            bases. Reads not uniquely
            placed in the genome and read pairs whose two ends were not
            consistently placed were
            discarded. After above filtering, NQS(30,25) standard was applied
            to all pairs of
            overlapping reads to call NQS bases and SNPs. Alignments (between
            two reads) with less
            than 100 NQS bases or with SNP rate > 0.01 were discarded. To
            exclude alignment between two
            copies of a single read, comparisons between two reads that share
            95% of their genome
            alignments (>=95% bases of read A and >=95% bases of read B were
            placed at the same locus
            of human genome) were discarded.
FEATURES             Location/Qualifiers
     source          1. .810
                     /organism="Pan troglodytes verus"
                     /mol_type="genomic DNA"
                     /sub_species="verus"
                     /db_xref="taxon:37012"
                     /clone_lib="Clint"
     STS             <1. .>810
ORIGIN

Alignment Scores:
Pred. No.:                191          Length:          810
Score:                    42.00        Matches:         6
Percent Similarity:       100.00%      Conservative:    0
Best Local Similarity:    100.00%      Mismatches:      0
Query Match:              100.00%      Indels:          0
DB:                       10           Gaps:            0

US-09-858-332G-28 (1-6) x BV527033 (1-810)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        405 CATAATCATAATCACAAT 388

RESULT 148
BC027815
LOCUS       BC027815                 820 bp    mRNA    linear   ROD 08-MAR-2005
DEFINITION  Mus musculus nerve growth factor receptor (TNFRSF16) associated
            protein 1, mRNA (cDNA clone MGC:36095 IMAGE:5250429), complete cds.
ACCESSION   BC027815
VERSION     BC027815.1  GI:20381288
KEYWORDS    MGC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 820)
  AUTHORS   Strausberg RL, Feingold EA, Grouse LH, Derge JG, Klausner RD,
            Collins FS, Wagner L, Shenmen CM, Schuler GD, Altschul SF, Zeeberg
            B, Buetow KH, Schaefer CF, Bhat NK, Hopkins RF, Jordan H, Moore T,
            Max SI, Wang J, Hsieh F, Diatchenko L, Marusina K, Farmer AA, Rubin
            GM, Hong L, Stapleton M, Soares MB, Bonaldo MF, Casavant TL,
            Scheetz TE, Brownstein MJ, Usdin TB, Toshiyuki S, Carninci P,
            Prange C, Raha SS, Loquellano NA, Peters GJ, Abramson RD, Mullahy
            SJ, Bosak SA, McEwan PJ, McKernan KJ, Malek JA, Gunaratne PH,
            Richards S, Worley KC, Hale S, Garcia AM, Gay LJ, Hulyk SW,
            Villalon DK, Muzny DM, Sodergren EJ, Lu X, Gibbs RA, Fahey J,
            Helton E, Ketteman M, Madan A, Rodrigues S, Sanchez A, Whiting M,
            Madan A, Young AC, Shevchenko Y, Bouffard GG, Blakesley RW,
            Touchman JW, Green ED, Dickson MC, Rodriguez AC, Grimwood J,
            Schmutz J, Myers RM, Butterfield YS, Krzywinski MI, Skalska U,
            Smailus DE, Schnerch A, Schein JE, Jones SJ and Marra MA.
  CONSRTM   Mammalian Gene Collection Program Team
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  PUBMED    12477932
REFERENCE   2  (bases 1 to 820)
  AUTHORS   Director MGC Project.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@nhgri.nih.gov
            Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
```

```
            Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D.,McCloskey,J.C.,
            McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
            Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
            Young,A., Zhang,L.-H. and Green,E.D.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 56 Row: b Column: 15
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 6753183.
FEATURES            Location/Qualifiers
     source         1. .820
                    /organism="Mus musculus"
                    /mol_type="mRNA"
                    /strain="mix FVB/N, C57BL/6J"
                    /db_xref="taxon:10090"
                    /clone="MGC:36095 IMAGE:5250429"
                    /tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months
                    old, gross tissue."
                    /clone_lib="NCI_CGAP_Mam5"
                    /lab_host="DH10B"
                    /note="Vector: pCMV-SPORT6"
     gene           1. .820
                    /gene="Ngfrap1"
                    /note="synonym: Nade"
                    /db_xref="GeneID:12070"
                    /db_xref="MGI:1338016"
     CDS            177. .551
                    /gene="Ngfrap1"
                    /codon_start=1
                    /product="nerve growth factor receptor (TNFRSF16)
                    associated protein 1"
                    /protein_id="AAH27815.1"
                    /db_xref="GI:20381289"
                    /db_xref="GeneID:12070"
                    /db_xref="MGI:1338016"
                    /translation="MANVHQENEEMEQPLQNGQEDRPVGGGEGHQPAANNNNNNHNHN
                    HNHHRRGQARRLAPNFRWAIPNRQMNDGLGGDGDDMEMFMEEMREIRRKLRELQLRNC
                    LRILMGELSNHHDHHDEFCLMP"
ORIGIN

Alignment Scores:
Pred. No.:                193             Length:         820
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       9               Gaps:           0

US-09-858-332G-28 (1-6) x BC027815 (1-820)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        297 CACAACCATAACCACAAC 314

RESULT 149
BV020456/c
LOCUS       BV020456                823 bp    DNA     linear   STS 30-MAY-2003
DEFINITION  S212P60204FH1.T0 CZECHII/Ei Mus musculus STS genomic, sequence
            tagged site.
ACCESSION   BV020456
VERSION     BV020456.1  GI:31104351
KEYWORDS    STS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 823)
  AUTHORS   Wade,C.M., Kulbokas,E.J. III, Kirby,A.W., Zody,M.C., Mullikin,J.C.,
            Lander,E.S., Lindblad-Toh,K. and Daly,M.J.
  TITLE     The mosaic structure of variation in the laboratory mouse genome
  JOURNAL   Nature 420 (6915), 574-578 (2002)
    PUBMED  12466852
COMMENT
            Contact: Kerstin Lindblad-Toh
            Whitehead Institute for Biomedical Research, Center for Genome
            Research
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172521477
            Fax: 6172580903
            Email: kersli@genome.wi.mit.edu
            Primer A: None
            Primer B: None
            STS size: 823
            Protocol:
            WGS-discovery: Paired-end low-coverage whole genome shotgun reads
            were generated from 129S1/SvImJ, C3H/HeJ, and BALB/cByJ. The WGS
            reads were placed uniquely on the MGSCv3 C57BL/6J assembly and SNP
            detection was carried out by SSAHA-SNP. 225,000 reads were
            annotated
            as STSs and 81,000 SNPs were annotated with alleles from C57BL/6J
            and the strain from which the particular read came. The validation
            rate for these SNPs was estimated at approximately 98%.
FEATURES            Location/Qualifiers
     source         1. .823
                    /organism="Mus musculus"
                    /mol_type="genomic DNA"
                    /strain="CZECHII/Ei"
                    /db_xref="taxon:10090"
                    /map="+ 9 25-743 40811431-40812149"
                    /clone_lib="CZECHII/Ei"
     STS            <1. .>823
ORIGIN

Alignment Scores:
Pred. No.:                194             Length:         823
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       10              Gaps:           0

US-09-858-332G-28 (1-6) x BV020456 (1-823)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        337 CATAATCATAACCACAAC 320

RESULT 150
BV618122
LOCUS       BV618122                837 bp    DNA     linear   STS 15-APR-2005
DEFINITION  S217P61862RD11.T0 Noemie Pan troglodytes troglodytes STS genomic,
            sequence tagged site.
ACCESSION   BV618122
VERSION     BV618122.1  GI:62608300
KEYWORDS    STS.
SOURCE      Pan troglodytes troglodytes
  ORGANISM  Pan troglodytes troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1  (bases 1 to 837)
  AUTHORS   Mikkelsen,T.S., Hillier,W.L., Eichler,E.E., Zody,M.C. and
            Jaffe,D.B.
  TITLE     Initial Sequence of the Chimpanzee Genome and Comparison with the
            Human Genome
  JOURNAL   Unpublished (2005)
COMMENT
            Contact: Michael C. Zody
            Broad Institute of MIT and Harvard
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172580933
            Fax: 6172580903
            Email: mczody@broad.mit.edu
            Primer A: No sequence submitted
```

Primer B: No sequence submitted
STS size: 837
Protocol:
23,021,928 chimpanzee whole genome shotgun reads were aligned to the Human genome NCBI
Build 34 (hg16,July 2003). Chimp WGS reads were from 9 donors, including Clint (Pan
troglodytes verus), 3 other Pan troglodytes verus chimps (Donald,Karlien,Yvonne), 3 Pan
troglodytes troglodytes chimps (Noemie,Masuku,Clara) and 2 chimps of unknown origin
(Gon,Unknown Chimp). Common names: Pan troglodytes verus is the western chimp and Pan
troglodytes troglodytes is the central chimp. To be included in chimpanzee SNP discovery, a
read must be at least 500bp in length, at least 50% of its base calls must have Phred
score >= 20, at least 30% of its base calls must satisfy SNQS(30,25)(single strand NQS, the
base in question has Phred score >= 30, the surrounding 10 bases in the read have Phred
score >= 25), and the read must have at least 200 bp SNQS(30,25) bases. Reads not uniquely
placed in the genome and read pairs whose two ends were not consistently placed were
discarded. After above filtering, NQS(30,25) standard was applied to all pairs of
overlapping reads to call NQS bases and SNPs. Alignments (between two reads) with less
than 100 NQS bases or with SNP rate > 0.01 were discarded. To exclude alignment between two
copies of a single read, comparisons between two reads that share 95% of their genome
alignments (>=95% bases of read A and >=95% bases of read B were placed at the same locus
of human genome) were discarded.

```
FEATURES             Location/Qualifiers
     source          1. .837
                     /organism="Pan troglodytes troglodytes"
                     /mol_type="genomic DNA"
                     /sub_species="troglodytes"
                     /db_xref="taxon:37011"
                     /clone_lib="Noemie"
     STS             <1. .>837
ORIGIN

Alignment Scores:
Pred. No.:              197          Length:        837
Score:                  42.00        Matches:       6
Percent Similarity:     100.00%      Conservative:  0
Best Local Similarity:  100.00%      Mismatches:    0
Query Match:            100.00%      Indels:        0
DB:                     10           Gaps:          0

US-09-858-332G-28 (1-6) x BV618122 (1-837)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         26 CACAATCATAACCACAAC 43

RESULT 151
AR550376
LOCUS       AR550376                852 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 5507 from patent US 6747137.
ACCESSION   AR550376
VERSION     AR550376.1  GI:53943551
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 852)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 5507 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .852
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              201          Length:        852
Score:                  42.00        Matches:       6
Percent Similarity:     100.00%      Conservative:  0
Best Local Similarity:  100.00%      Mismatches:    0
Query Match:            100.00%      Indels:        0
DB:                     6            Gaps:          0

US-09-858-332G-28 (1-6) x AR550376 (1-852)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        315 CACAACCACAACCACAAC 332

RESULT 152
AF097440
LOCUS       AF097440                854 bp    mRNA    linear   ROD 13-APR-1999
DEFINITION  Mus musculus brain expressed X-linked protein 3 (Bex3) mRNA,
            complete cds.
ACCESSION   AF097440
VERSION     AF097440.1  GI:4580593
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 854)
  AUTHORS   Brown,A.L. and Kay,G.F.
  TITLE     Bex1, a gene with increased expression in parthenogenetic embryos,
            is a member of a novel gene family on the mouse X chromosome
  JOURNAL   Hum. Mol. Genet. 8 (4), 611-619 (1999)
   PUBMED   10072429
REFERENCE   2  (bases 1 to 854)
  AUTHORS   Brown,A.L. and Kay,G.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-1998) Cancer Unit, Queensland Institute of
            Medical Research, Herston Rd, Brisbane, Qld 4029, Australia
FEATURES             Location/Qualifiers
     source          1. .854
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="dbEST:AA272375"
                     /db_xref="taxon:10090"
                     /chromosome="X"
                     /map="near Plp"
                     /tissue_type="pooled organs"
     gene            1. .854
                     /gene="Bex3"
     CDS             172. .546
                     /gene="Bex3"
                     /codon_start=1
                     /product="brain expressed X-linked protein 3"
                     /protein_id="AAD24431.1"
                     /db_xref="GI:4580594"
                     /translation="MANVHQENEEMEQPLQNGQEDRPVGGGEGHQPAANNNNNNHNHN
                     HNHHRRGQARRLAPNFRWAIPNRQMNDGLGGDGDDMEMFMEEMREIRRKLRELQLRNC
                     LRILMGELSNHHDHHDEFCLMP"
ORIGIN

Alignment Scores:
Pred. No.:              201          Length:        854
Score:                  42.00        Matches:       6
```

```
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     9              Gaps:          0

US-09-858-332G-28 (1-6) x AF097440 (1-854)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        292 CACAACCATAACCACAAC 309

RESULT 153
AR501099
LOCUS       AR501099                858 bp    DNA     linear   PAT 22-SEP-2004
DEFINITION  Sequence 6059 from patent US 6703491.
ACCESSION   AR501099
VERSION     AR501099.1  GI:52436574
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 858)
  AUTHORS   Homburger,S.A., Ebens,A.J. Jr., Erickson,C.S., Francis-Lang,H.L.,
            Margolis,J.S., Reddy,B.P., Ruddy,D.A. and Buchman,A.R.
  TITLE     Drosophila sequences
  JOURNAL   Patent: US 6703491-A 6059 09-MAR-2004;
            Exelixis, Inc.; South San Francisco, CA
FEATURES             Location/Qualifiers
     source          1. .858
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              202            Length:        858
Score:                  42.00          Matches:       6
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     6              Gaps:          0

US-09-858-332G-28 (1-6) x AR501099 (1-858)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        140 CATAACCATAATCATAAC 157

RESULT 154
AR516381
LOCUS       AR516381                858 bp    DNA     linear   PAT 22-SEP-2004
DEFINITION  Sequence 21341 from patent US 6703491.
ACCESSION   AR516381
VERSION     AR516381.1  GI:52451856
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 858)
  AUTHORS   Homburger,S.A., Ebens,A.J. Jr., Erickson,C.S., Francis-Lang,H.L.,
            Margolis,J.S., Reddy,B.P., Ruddy,D.A. and Buchman,A.R.
  TITLE     Drosophila sequences
  JOURNAL   Patent: US 6703491-A 21341 09-MAR-2004;
            Exelixis, Inc.; South San Francisco, CA
FEATURES             Location/Qualifiers
     source          1. .858
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              202            Length:        858
Score:                  42.00          Matches:       6
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     6              Gaps:          0

US-09-858-332G-28 (1-6) x AR516381 (1-858)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        140 CATAACCATAATCATAAC 157

RESULT 155
AR549587
LOCUS       AR549587                858 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 4718 from patent US 6747137.
ACCESSION   AR549587
VERSION     AR549587.1  GI:53942762
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 858)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 4718 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .858
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              202            Length:        858
Score:                  42.00          Matches:       6
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     6              Gaps:          0

US-09-858-332G-28 (1-6) x AR549587 (1-858)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        506 CACAACCACAACCACAAC 523

RESULT 156
BV518914/c
LOCUS       BV518914                866 bp    DNA     linear   STS 08-APR-2005
DEFINITION  G591P69286FF2.T0 Clint Pan troglodytes verus STS genomic, sequence
            tagged site.
ACCESSION   BV518914
VERSION     BV518914.1  GI:62396672
KEYWORDS    STS.
SOURCE      Pan troglodytes verus
  ORGANISM  Pan troglodytes verus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1  (bases 1 to 866)
  AUTHORS   Mikkelsen,T.S., Hillier,W.L., Eichler,E.E., Zody,M.C. and
            Jaffe,D.B.
  TITLE     Initial Sequence of the Chimpanzee Genome and Comparison with the
            Human Genome
  JOURNAL   Unpublished (2005)
COMMENT
            Contact: Michael C. Zody
            Broad Institute of MIT and Harvard
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172580933
            Fax: 6172580903
            Email: mczody@broad.mit.edu
            Primer A: No sequence submitted
```

```
            Primer B: No sequence submitted
            STS size: 866
            Protocol:
            23,021,928 chimpanzee whole genome shotgun reads were aligned to
            the Human genome NCBI
            Build 34 (hg16,July 2003). Chimp WGS reads were from 9 donors,
            including Clint (Pan
            troglodytes verus), 3 other Pan troglodytes verus chimps
            (Donald,Karlien,Yvonne), 3 Pan
            troglodytes troglodytes chimps (Noemie,Masuku,Clara) and 2 chimps
            of unknown origin
            (Gon,Unknown Chimp). Common names:  Pan troglodytes verus is the
            western chimp and Pan
            troglodytes troglodytes is the central chimp. To be included in
            chimpanzee SNP discovery, a
            read must be at least 500bp in length, at least 50% of its base
            calls must have Phred
            score >= 20, at least 30% of its base calls must satisfy
            SNQS(30,25)(single strand NQS, the
            base in question has Phred score >= 30, the surrounding 10 bases in
            the read have Phred
            score >= 25), and the read must have at least 200 bp SNQS(30,25)
            bases. Reads not uniquely
            placed in the genome and read pairs whose two ends were not
            consistently placed were
            discarded. After above filtering, NQS(30,25) standard was applied
            to all pairs of
            overlapping reads to call NQS bases and SNPs. Alignments (between
            two reads) with less
            than 100 NQS bases or with SNP rate > 0.01 were discarded. To
            exclude alignment between two
            copies of a single read, comparisons between two reads that share
            95% of their genome
            alignments (>=95% bases of read A and >=95% bases of read B were
            placed at the same locus
            of human genome) were discarded.
FEATURES             Location/Qualifiers
     source          1. .866
                     /organism="Pan troglodytes verus"
                     /mol_type="genomic DNA"
                     /sub_species="verus"
                     /db_xref="taxon:37012"
                     /clone_lib="Clint"
     STS             <1. .>866
ORIGIN

Alignment Scores:
Pred. No.:              204             Length:         866
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     10              Gaps:           0

US-09-858-332G-28 (1-6) x BV518914 (1-866)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        216 CATAATCATAATCATAAT 199

RESULT 157
BT020557
LOCUS       BT020557                867 bp    mRNA    linear   PLN 30-JAN-2005
DEFINITION  Arabidopsis thaliana At5g49920 gene, complete cds.
ACCESSION   BT020557
VERSION     BT020557.1  GI:58331810
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 867)
  AUTHORS   Kim,C.J., Chen,H., Cheuk,R., Shinn,P. and Ecker,J.R.
  TITLE     Arabidopsis ORF clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 867)
  AUTHORS   Kim,C.J., Chen,H., Cheuk,R., Shinn,P. and Ecker,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JAN-2005) Salk Institute Genomic Analysis Laboratory
            (SIGnAL), Plant Biology Laboratory, The Salk Institute for
            Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
            USA
FEATURES             Location/Qualifiers
     source          1. .867
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /chromosome="5"
                     /clone="U69229"
                     /ecotype="Columbia"
                     /note="This clone is in pUNI 51"
     CDS             1. .867
                     /note="unknown protein"
                     /codon_start=1
                     /product="At5g49920"
                     /protein_id="AAW70403.1"
                     /db_xref="GI:58331811"
                     /translation="MAEERSPSKVKFMCSFGGRILPRPSDSVLKYVGGETRVVAVSPD
                     ISFSELMKKLTAITENDIVLKYQIIPEDLDALVSVKSDEDVKHMIEEYNRHETPKLRT
                     FLFPANPVVLESQLGPIEPQTIEQRYIEAINGILRTSKSATALRAPIKTRPSFTVSTC
                     SSPKSESPDGYSHEPPETSFQNNYQLSRLYPMHRVHSSPNISHQLQPHSHYHNHNHNA
                     YVQPPTNYLTCRPRLPTPLEIPRGIGWDQSYTTTHNSGGGNGKYGCNDDRRFWGRASS
                     VPQSPRNHGLRL"
ORIGIN

Alignment Scores:
Pred. No.:              204             Length:         867
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x BT020557 (1-867)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        634 CACAACCACAACCACAAT 651

RESULT 158
AY143847
LOCUS       AY143847                882 bp    mRNA    linear   PLN 05-OCT-2002
DEFINITION  Arabidopsis thaliana At3g12480/MQC3.32 mRNA, complete cds.
ACCESSION   AY143847
VERSION     AY143847.1  GI:23505852
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 882)
  AUTHORS   Cheuk,R., Chen,H., Kim,C.J., Shinn,P., Bowser,L., Carninci,P.,
            Chan,M.M., Chang,C.H., Dale,J.M., Deng,J.M., Hayashizaki,Y.,
            Hsuan,V.W., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G.,
            Kawai,J., Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M.,
            Nguyen,M., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M.,
            Southwick,A., Tang,C.C., Toriumi,M., Wallender,E.K., Wong,C.,
            Wu,H.C., Yamada,K., Yu,G., Yuan,S., Shinozaki,K., Davis,R.W.,
            Theologis,A. and Ecker,J.R.
  TITLE     Arabidopsis ORF clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 882)
  AUTHORS   Cheuk,R., Chen,H., Kim,C.J., Shinn,P., Bowser,L., Carninci,P.,
            Chan,M.M., Chang,C.H., Dale,J.M., Deng,J.M., Hayashizaki,Y.,
```

Hsuan,V.W., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A., Tang,C.C., Toriumi,M., Wallender,E.K., Wong,C., Wu,H.C., Yamada,K., Yu,G., Yuan,S., Shinozaki,K., Davis,R.W., Theologis,A. and Ecker,J.R.

TITLE Direct Submission

JOURNAL Submitted (19-AUG-2002) Salk Institute Genomic Analysis Laboratory (SIGnAL), Plant Biology Laboratory, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

COMMENT RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN Arabidopsis Full-Length cDNA') : Seki,M., Narusaka,M., Ishida,J., Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J., Hayashizaki,Y. and Shinozaki,K.

The Salk, Stanford, PGEC (SSP) Consortium members constructed and sequenced the pUNI (ORF) clones using the RAFL cDNAs: Cheuk,R., Chen,H., Kim,C.J., Shinn,P., Banh,J. Bowser,L., Chan,M.M., Chang,E., Dale,J.M., Deng, J.M., Goldsmith,A.D., Jones,T., Karlin-Neumann,G., Lam,B., Lee,J.M., Lin,J., Miranda,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C., Yamada,K., Yamamura, Y., Yu,G., Davis,R.W., Theologis,A., and Ecker,J.R.

Cheuk,R. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to this work. Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk) contributed equally to this work as PIs.

FEATURES Location/Qualifiers

source 1. .882
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="3"
/clone="U21457"
/ecotype="Columbia"
/note="This clone is in pUNI 51"

CDS 1. .882
/note="unknown protein"
/codon_start=1
/product="At3g12480/MQC3.32"
/protein_id="AAN28786.1"
/db_xref="GI:23505853"
/translation="MRKKLDTRFPAARIKKIMQADEDVGKIALAVPVLVSKSLELFLQ
DLCDRTYEITLERGAKTVSSLHLKHCVERYNVFDFLREVVSKVPDYGHSQGQGHGDVT
MDDRSISKRRKPISDEVNDSDEEYKKSKTQEIGSAKTSGRGGRGRGRGRGGRAAKA
AEREGLNREMEVEAANSGQPPPEDNVKMHASESSPQEDEKKGIDGTAASNEDTKQHLQ
SPKEGIDFDLNAESLDLNETKLAPATGTTTTTAATDSEEYSGWPMMDISKMDPAQLA
SLGKRIDEDEEDYDEEG"

ORIGIN

Alignment Scores:

Pred. No.:	207	Length:	882
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	15	Gaps:	0

US-09-858-332G-28 (1-6) x AY143847 (1-882)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        735 CACAACCACAACCACAAC 752
```

RESULT 159
BC058503
LOCUS BC058503 887 bp mRNA linear ROD 05-APR-2005
DEFINITION Rattus norvegicus nerve growth factor receptor (TNFRSF16) associated protein 1, mRNA (cDNA clone MGC:72988 IMAGE:6887969), complete cds.
ACCESSION BC058503
VERSION BC058503.1 GI:34849868
KEYWORDS MGC.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 887)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.

CONSRTM Mammalian Gene Collection Program Team
TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBMED 12477932

REFERENCE 2 (bases 1 to 887)
AUTHORS .
CONSRTM NIH MGC Project
TITLE Direct Submission
JOURNAL Submitted (16-SEP-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov

COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: John C. Marshall, M.D., Ph.D
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca
Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield, Andy Chan, Steve S. Chand, William Chow, Alison Cloutier, Ruth Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao, Kim MacDonald, Amara Masson, Mike R. Mayo, Josh Moran, Ryan Morin, Teika Olson, Diana Palmquist, Anca Petrescu, Anna Liisa Prahbu, Parvaneh Saeedi, JR Santos, Angelique Schnerch, Ursula Skalska, Duane Smailus, Jeff Stott, Miranda Tsai, George Yang, Jacquie Schein, Asim Siddiqui, Rob Holt, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 52 Row: c Column: 1
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein.

FEATURES Location/Qualifiers

source 1. .887
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="MGC:72988 IMAGE:6887969"
/tissue_type="Pituitary gland, anterior, rat"
/clone_lib="NICHD_Rr_Pit1"
/lab_host="DH10B"
/note="Vector: pDNR-LIB"

gene 1. .887
/gene="Ngfrap1"

```
                  /note="synonyms: Bex3, Nade"
                  /db_xref="GeneID:117089"
                  /db_xref="RGD:3148"
     CDS          218. .610
                  /gene="Ngfrap1"
                  /codon_start=1
                  /product="Ngfrap1 protein"
                  /protein_id="AAH58503.1"
                  /db_xref="GI:34849869"
                  /db_xref="GeneID:117089"
                  /db_xref="RGD:3148"
                  /translation="MANIHQENEEMEQPLQNGQEDRPVGGGEGHQPAANNNNHNHNHN
                  HNHNHNHNHHRRGQARRLAPNFRWAIPNRQMNDGLGGDGDDMEMFMEEMREIRRKLRE
                  LQLRNCLRILMGELSNHHDHHDEFCLMP"
ORIGIN

Alignment Scores:
Pred. No.:              208           Length:        887
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     9             Gaps:          0

US-09-858-332G-28 (1-6) x BC058503 (1-887)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        332 CACAACCATAACCACAAC 349

RESULT 160
AR548322
LOCUS       AR548322                888 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 3453 from patent US 6747137.
ACCESSION   AR548322
VERSION     AR548322.1  GI:53941497
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 888)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 3453 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .888
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              209           Length:        888
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-28 (1-6) x AR548322 (1-888)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        173 CACAACCACAACCACAAT 190

RESULT 161
AR549211
LOCUS       AR549211                903 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 4342 from patent US 6747137.
ACCESSION   AR549211
VERSION     AR549211.1  GI:53942386
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 903)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 4342 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .903
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              212           Length:        903
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-28 (1-6) x AR549211 (1-903)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        308 CACAACCACAACCACAAC 325

RESULT 162
A57713
LOCUS       A57713                  912 bp    DNA     linear   PAT 03-MAR-1998
DEFINITION  Sequence 23 from Patent WO9634103.
ACCESSION   A57713
VERSION     A57713.1  GI:3713537
KEYWORDS    .
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified sequences.
REFERENCE   1
  AUTHORS   Hamers,R. and Muyldermans,S.
  TITLE     Variable fragments of immunoglobulins - use for therapeutic or
            veterinary purposes
  JOURNAL   Patent: WO 9634103-A 23 31-OCT-1996;
            UNIV BRUXELLES (BE)
COMMENT     Other publication AU 5647896 961118.
FEATURES             Location/Qualifiers
     source          1. .912
                     /organism="unidentified"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32644"
     CDS             1. .912
                     /note="unnamed protein product; Protein sequence is in
                     conflict with the conceptual translation"
                     /codon_start=1
                     /protein_id="CAA03460.1"
                     /db_xref="GI:3713538"
                     /translation="DVQLQASGGGSVQAGGSLRLSCAASGYTIGPYCMGWFRQAPGKE
                     REGVAAINMGGGITYYADSVKGRFTISQDNAKNTVYLLMNSLEPEDTAIYYCAADSTI
                     YASYYECGHGLSTGGYGYDSWGQGTQVTVSSEPKIPQPQPKPQPQPQPQPKPQPKPEP
                     MAEVQLQASGGGSVQAGQSLRLSCATSGATSSSNCMGWFRQAPGKEREGVAVIDTGRG
                     NTAYADSVQGRLTISLDNAKNTLYLQMNSLKPEDTAMYYCAADTSTWYRGYCGTNPNY
                     FSYWGQGTQVTVSSGRYDVPDYGSGRA"
ORIGIN

Alignment Scores:
Pred. No.:              214           Length:        912
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     6             Gaps:          0
```

```
US-09-858-332G-28 (1-6) x A57713 (1-912)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        431 CACAACCACAACCACAAC 448

RESULT 163
AR551978
LOCUS       AR551978                921 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 7109 from patent US 6747137.
ACCESSION   AR551978
VERSION     AR551978.1  GI:53945153
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 921)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 7109 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .921
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              216             Length:         921
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x AR551978 (1-921)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        478 CATAATCATAACCACAAT 495

RESULT 164
CQ610476
LOCUS       CQ610476                927 bp    DNA     linear   PAT 02-FEB-2004
DEFINITION  Sequence 38234 from Patent WO0171042.
ACCESSION   CQ610476
VERSION     CQ610476.1  GI:41662297
KEYWORDS    .
SOURCE      Drosophila sp.
  ORGANISM  Drosophila sp.
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1
  AUTHORS   Venter,J.C., Adams,M., Li,P.W. and Myers,E.W.
  TITLE     Detection kits, such as nucleic acid arrays, for detecting the
            expression of 10,000 or more Drosophila genes and uses thereof
  JOURNAL   Patent: WO 0171042-A 38234 27-SEP-2001;
            PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1. .927
                     /organism="Drosophila sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:7242"
ORIGIN

Alignment Scores:
Pred. No.:              217             Length:         927
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x CQ610476 (1-927)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        629 CACAACCACAACCACAAT 646

RESULT 165
AR545979
LOCUS       AR545979                966 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 1110 from patent US 6747137.
ACCESSION   AR545979
VERSION     AR545979.1  GI:53939154
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 966)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 1110 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .966
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              226             Length:         966
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x AR545979 (1-966)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        659 CACAATCACAATCACAAT 676

RESULT 166
CNS06HVC
LOCUS       CNS06HVC                976 bp    DNA     linear   STS 30-NOV-2001
DEFINITION  T7 end of clone AS0AA015E09 of library AS0AA from strain CLIB 533
            of Saccharomyces bayanus, sequence tagged site.
ACCESSION   AL399502
VERSION     AL399502.1  GI:12154043
KEYWORDS    STS.
SOURCE      Saccharomyces bayanus
  ORGANISM  Saccharomyces bayanus
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 976)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  PUBMED    11152876
REFERENCE   2  (bases 1 to 976)
  AUTHORS   Bon,E., Neuveglise,C., Casaregola,S., Artiguenave,F., Wincker,P.,
            Aigle,M. and Durrens,P.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 5.
            Saccharomyces bayanus var. uvarum
  JOURNAL   FEBS Lett. 487 (1), 37-41 (2000)
```

```
   PUBMED   11152880
REFERENCE   3  (bases 1 to 976)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This STS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .976
                     /organism="Saccharomyces bayanus"
                     /mol_type="genomic DNA"
                     /strain="CLIB 533"
                     /variety="uvarum"
                     /db_xref="taxon:4931"
                     /clone="AS0AA015E09"
                     /clone_lib="AS0AA"
                     /note="end : T7"
     misc_feature    <515. .>975
                     /note="similar to Saccharomyces cerevisiae ORF YPL229w [
                     weak similarity to YMR181c ]
                     1 putative frameshift(s)"
                     /evidence=not_experimental
ORIGIN

Alignment Scores:
Pred. No.:              229           Length:        976
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     10            Gaps:          0

US-09-858-332G-28 (1-6) x CNS06HVC (1-976)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        651 CACAACCACAACCACAAC 668

RESULT 167
CNS06HW9/c
LOCUS       CNS06HW9                  994 bp    DNA     linear   STS 30-NOV-2001
DEFINITION  T3 end of clone AS0AA015G03 of library AS0AA from strain CLIB 533
            of Saccharomyces bayanus, sequence tagged site.
ACCESSION   AL399535
VERSION     AL399535.1  GI:12154132
KEYWORDS    STS.
SOURCE      Saccharomyces bayanus
  ORGANISM  Saccharomyces bayanus
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 994)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
   PUBMED   11152876
REFERENCE   2  (bases 1 to 994)
  AUTHORS   Bon,E., Neuveglise,C., Casaregola,S., Artiguenave,F., Wincker,P.,
            Aigle,M. and Durrens,P.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 5.
            Saccharomyces bayanus var. uvarum
  JOURNAL   FEBS Lett. 487 (1), 37-41 (2000)
   PUBMED   11152880
REFERENCE   3  (bases 1 to 994)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This STS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .994
                     /organism="Saccharomyces bayanus"
                     /mol_type="genomic DNA"
                     /strain="CLIB 533"
                     /variety="uvarum"
                     /db_xref="taxon:4931"
                     /clone="AS0AA015G03"
                     /clone_lib="AS0AA"
                     /note="end : T3"
     STS             1. .994
ORIGIN

Alignment Scores:
Pred. No.:              233           Length:        994
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     10            Gaps:          0

US-09-858-332G-28 (1-6) x CNS06HW9 (1-994)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        444 CATAACCATAACCATAAC 427

RESULT 168
CNS06EEE/c
LOCUS       CNS06EEE                 1022 bp    DNA     linear   STS 10-JAN-2001
DEFINITION  T7 end of clone AR0AA020H09 of library AR0AA from strain CBS 732 of
            Zygosaccharomyces rouxii, sequence tagged site.
ACCESSION   AL395004
VERSION     AL395004.1  GI:12146304
KEYWORDS    STS.
SOURCE      Zygosaccharomyces rouxii
  ORGANISM  Zygosaccharomyces rouxii
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Zygosaccharomyces.
REFERENCE   1  (bases 1 to 1022)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
   PUBMED   11152876
REFERENCE   2  (bases 1 to 1022)
  AUTHORS   de Montigny,J., Straub,M., Potier,S., Tekaia,F., Dujon,B.,
```

```
            Wincker,P., Artiguenave,F. and Souciet,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 8.
            Zygosaccharomyces rouxii
  JOURNAL   FEBS Lett. 487 (1), 52-55 (2000)
  PUBMED    11152883
REFERENCE   3  (bases 1 to 1022)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This STS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .1022
                     /organism="Zygosaccharomyces rouxii"
                     /mol_type="genomic DNA"
                     /strain="CBS 732"
                     /db_xref="taxon:4956"
                     /clone="AR0AA020H09"
                     /clone_lib="AR0AA"
                     /note="end : T7"
     misc_feature    complement(<242. .>846)
                     /note="similar to Saccharomyces cerevisiae ORF YNL167c [
                     SKO1 ; cre-binding bzip protein ]
                     1 putative frameshift(s)"
                     /evidence=not_experimental
ORIGIN

Alignment Scores:
Pred. No.:                239           Length:          1022
Score:                    42.00         Matches:         6
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       10            Gaps:            0

US-09-858-332G-28 (1-6) x CNS06EEE (1-1022)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        670 CATAATCACAATCATAAT 653

RESULT 169
BT020324
LOCUS       BT020324               1027 bp    mRNA    linear   PLN 04-DEC-2004
DEFINITION  Arabidopsis thaliana At5g49920 gene, complete cds.
ACCESSION   BT020324
VERSION     BT020324.1  GI:56381920
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1027)
  AUTHORS   Kim,C.J., Chen,H., Cheuk,R., Shinn,P. and Ecker,J.R.
  TITLE     Arabidopsis ORF clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1027)
  AUTHORS   Kim,C.J., Chen,H., Cheuk,R., Shinn,P. and Ecker,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2004) Salk Institute Genomic Analysis Laboratory
            (SIGnAL), Plant Biology Laboratory, The Salk Institute for
            Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
            USA
FEATURES             Location/Qualifiers
     source          1. .1027
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /chromosome="5"
                     /clone="S69229"
                     /ecotype="Columbia"
                     /note="This clone is in pUNI 51"
     CDS             65. .931
                     /note="unknown protein"
                     /codon_start=1
                     /product="At5g49920"
                     /protein_id="AAV85679.1"
                     /db_xref="GI:56381921"
                     /translation="MAEERSPSKVKFMCSFGGRILPRPSDSVLKYVGGETRVVAVSPD
                     ISFSELMKKLTAITENDIVLKYQIIPEDLDALVSVKSDEDVKHMIEEYNRHETPKLRT
                     FLFPANPVVLESQLGPIEPQTIEQRYIEAINGILRTSKSATALRAPIKTRPSFTVSTC
                     SSPKSESPDGYSHEPPETSFQNNYQLSRLYPMHRVHSSPNISHQLQPHSHYHNHNHNA
                     YVQPPTNYLTCRPRLPTPLEIPRGIGWDQSYTTTHNSGGGNGKYGCNDDRRFWGRASS
                     VPQSPRNHGLRL"
ORIGIN

Alignment Scores:
Pred. No.:                240           Length:          1027
Score:                    42.00         Matches:         6
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       15            Gaps:            0

US-09-858-332G-28 (1-6) x BT020324 (1-1027)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        698 CACAACCACAACCACAAT 715

RESULT 170
AR545582
LOCUS       AR545582               1035 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 713 from patent US 6747137.
ACCESSION   AR545582
VERSION     AR545582.1  GI:53938757
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1035)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 713 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .1035
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:                242           Length:          1035
Score:                    42.00         Matches:         6
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       6             Gaps:            0

US-09-858-332G-28 (1-6) x AR545582 (1-1035)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        857 CACAATCACAATCACAAT 874
```

```
RESULT 171
AR545202
LOCUS       AR545202               1065 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 333 from patent US 6747137.
ACCESSION   AR545202
VERSION     AR545202.1  GI:53938377
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1065)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 333 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .1065
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              249           Length:        1065
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-28 (1-6) x AR545202 (1-1065)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        596 CACAACCACAACCACAAC 613

RESULT 172
AF457874/c
LOCUS       AF457874               1087 bp    mRNA    linear   PLN 02-JAN-2003
DEFINITION  Euphorbia serrata ribosome inactivating protein Euserratin 1
            precursor (Eus1) mRNA, complete cds.
ACCESSION   AF457874
VERSION     AF457874.1  GI:27462646
KEYWORDS    .
SOURCE      Euphorbia serrata
  ORGANISM  Euphorbia serrata
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Malpighiales; Euphorbiaceae; Euphorbioideae;
            Euphorbieae; Euphorbia.
REFERENCE   1  (bases 1 to 1087)
  AUTHORS   Girbes,T., Arias,F.J. and Benvenuto,E.
  TITLE     Purification, characterization and molecular cloning of
            euserratins, new type 1 ribosome-inactivating proteins from
            Euphorbia serrata L.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1087)
  AUTHORS   Girbes,T., Arias,F.J. and Benvenuto,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-DEC-2001) Bioquimica, Biologia Molecular y
            Fisiologia, Universidad de Valladolid, Facultad de Ciencias. Dr.
            Mergelina s/n, Valladolid 47005, Spain
FEATURES             Location/Qualifiers
     source          1. .1087
                     /organism="Euphorbia serrata"
                     /mol_type="mRNA"
                     /db_xref="taxon:196589"
     gene            1. .1087
                     /gene="Eus1"
     5'UTR           1. .64
                     /gene="Eus1"
     CDS             65. .958
                     /gene="Eus1"
                     /EC_number="3.2.2.22"
                     /function="depurinates 28S rRNA in ribosomes"
                     /function="inhibits protein synthesis"
                     /note="type 1 ribosome inactivating protein;
                     N-glycosidase"
                     /codon_start=1
                     /product="ribosome inactivating protein Euserratin 1
                     precursor"
                     /protein_id="AAO15530.1"
                     /db_xref="GI:27462647"
                     /translation="MKMWIVFATWLCCNIAIGSSLNAPVNYPTVKFTTHLASVGSYQT
                     LMSALRVNLESKLESHNIPLLRKPSDITDQNKYLLVELTNYDTKRTVTLALTVLNVYV
                     IGYKSGTKSFFLKDAPSDAKTLLFTDTTPKTLEVDTNYNNLGDRSKVGLGIPALKNAI
                     NILNQFDGVSTDQDFKHSLLIVIQMVSEAARFKFIQLKIEGGLLTQYLPKPDTISYQN
                     NWSALSKSIQLADANGRLSESVTLKYEDGKDRVVFMVEQVQRDISLLLYHVGSTERLG
                     EEKFGGVYADEAVEDELLSVI"
     sig_peptide     65. .130
                     /gene="Eus1"
     mat_peptide     131. .955
                     /gene="Eus1"
                     /product="ribosome inactivating protein Euserratin 1"
     3'UTR           959. .1087
                     /gene="Eus1"
ORIGIN

Alignment Scores:
Pred. No.:              254           Length:        1087
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     15            Gaps:          0

US-09-858-332G-28 (1-6) x AF457874 (1-1087)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1010 CACAATCACAATCACAAT 993

RESULT 173
AK106177
LOCUS       AK106177               1095 bp    mRNA    linear   PLN 24-JUL-2003
DEFINITION  Oryza sativa (japonica cultivar-group) cDNA clone:001-208-C09, full
            insert sequence.
ACCESSION   AK106177
VERSION     AK106177.1  GI:32991386
KEYWORDS    FLI_CDNA; oligo capping.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   The Rice Full-Length cDNA Consortium, National Institute of
            Agrobiological Sciences Rice Full-Length cDNA Project Team:,
            Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K.,
            Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I.,
            Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C.,
            Ohtsuki,K., Shishiki,T., Foundation of Advancement of International
            Science Genome Sequencing & Analysis Group:, Otomo,Y., Murakami,K.,
            Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y.,
            Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M.,
            Narikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Niikura,J.,
            Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Miura,J.,
            Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN:,
            Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuda,S.,
            Hara,A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M.,
            Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Osato,N., Ota,Y.,
            Saito,R., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
            Yoshino,M. and Hayashizaki,Y.
  TITLE     Collection, mapping, and annotation of over 28,000 cDNA clones from
            japonica rice
```

JOURNAL Science 301 (5631), 376-379 (2003)
PUBMED 12869764
REFERENCE 2 (bases 1 to 1095)
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K., Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K., Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawagashira,N., Kawai,J., Kawamata,M., Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M., Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Kurosaki,T., Kusumegi,T., Li,C., Lu,M., Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A., Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakamura,M., Namiki,T., Narikawa,R., Niikura,J., Nishi,K., Nomura,K., Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M., Ooka,H., Osato,N., Ota,Y., Otomo,Y., Ryu,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Shibata,K., Shinagawa,A., Shiraki,T., Shishiki,T., Sogabe,Y., Sugano,S., Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,W., Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and Yoshimura,A.
TITLE Direct Submission
JOURNAL Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)
COMMENT This clone is one of the 28K full-length cDNA clones from japonica rice.
URL : http://cdna01.dna.affrc.go.jp/cDNA/
NIAS Rice Full-Length cDNA Project Team: Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T. and Yamamoto,M.
FAIS Genome Sequencing & Analysis Group: Otomo,Y., Iida,Y., Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M., Kobayashi,M., Kodama,T., Kurosaki,T., Kusumegi,T., Lu,M., Masuda,H., Miura,J., Mizuno,K., Narikawa,R., Niikura,J., Oka,M., Ryu,R., Sugano,S., Sugiyama,A., Suzuki,Y., Tsunoda,Y., Ueda,M., Xie,Q., Yokomizo,S., Yoshimura,A., Matsubara,K. and Murakami,K.
Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Iida,J., Imamura,K., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J., Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Osato,N., Ota,Y., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K., Yasunishi,A. and Hayashizaki,Y.

FEATURES Location/Qualifiers
source 1. .1095
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="001-208-C09"

ORIGIN

Alignment Scores:

Pred. No.:	255	Length:	1095
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0

DB: 15 Gaps: 0

US-09-858-332G-28 (1-6) x AK106177 (1-1095)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        312 CACAACCACAACCACAAC 329
```

RESULT 174
AK104498
LOCUS AK104498 1105 bp mRNA linear PLN 24-JUL-2003
DEFINITION Oryza sativa (japonica cultivar-group) cDNA clone:006-302-D10, full insert sequence.
ACCESSION AK104498
VERSION AK104498.1 GI:32989707
KEYWORDS FLI_CDNA; oligo capping.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1
AUTHORS The Rice Full-Length cDNA Consortium, National Institute of Agrobiological Sciences Rice Full-Length cDNA Project Team:, Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group:, Otomo,Y., Murakami,K., Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y., Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M., Narikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Niikura,J., Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Miura,J., Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN:, Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Osato,N., Ota,Y., Saito,R., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Yoshino,M. and Hayashizaki,Y.
TITLE Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice
JOURNAL Science 301 (5631), 376-379 (2003)
PUBMED 12869764
REFERENCE 2 (bases 1 to 1105)
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K., Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K., Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawagashira,N., Kawai,J., Kawamata,M., Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M., Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Kurosaki,T., Kusumegi,T., Li,C., Lu,M., Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A., Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakamura,M., Namiki,T., Narikawa,R., Niikura,J., Nishi,K., Nomura,K., Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M., Ooka,H., Osato,N., Ota,Y., Otomo,Y., Ryu,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Shibata,K., Shinagawa,A., Shiraki,T., Shishiki,T., Sogabe,Y., Sugano,S., Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,W., Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and Yoshimura,A.
TITLE Direct Submission
JOURNAL Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)
COMMENT This clone is one of the 28K full-length cDNA clones from japonica rice.

URL : http://cdna01.dna.affrc.go.jp/cDNA/
NIAS Rice Full-Length cDNA Project Team: Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T. and Yamamoto,M.
FAIS Genome Sequencing & Analysis Group: Otomo,Y., Iida,Y., Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M., Kobayashi,M., Kodama,T., Kurosaki,T., Kusumegi,T., Lu,M., Masuda,H., Miura,J., Mizuno,K., Narikawa,R., Niikura,J., Oka,M., Ryu,R., Sugano,S., Sugiyama,A., Suzuki,Y., Tsunoda,Y., Ueda,M., Xie,Q., Yokomizo,S., Yoshimura,A., Matsubara,K. and Murakami,K.
Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Iida,J., Imamura,K., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J., Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Osato,N., Ota,Y., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K., Yasunishi,A. and Hayashizaki,Y.

FEATURES Location/Qualifiers
source 1. .1105
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="006-302-D10"

ORIGIN

Alignment Scores:

Pred. No.:	258	Length:	1105
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	15	Gaps:	0

US-09-858-332G-28 (1-6) x AK104498 (1-1105)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        322 CACAACCACAACCACAAC 339
```

RESULT 175
AK104803
LOCUS AK104803 1105 bp mRNA linear PLN 24-JUL-2003
DEFINITION Oryza sativa (japonica cultivar-group) cDNA clone:001-040-C08, full insert sequence.
ACCESSION AK104803
VERSION AK104803.1 GI:32990012
KEYWORDS FLI_CDNA; oligo capping.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1
AUTHORS The Rice Full-Length cDNA Consortium, National Institute of Agrobiological Sciences Rice Full-Length cDNA Project Team:, Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group:, Otomo,Y., Murakami,K., Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y., Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M., Narikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Niikura,J., Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Miura,J., Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN:, Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Osato,N., Ota,Y., Saito,R., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Yoshino,M. and Hayashizaki,Y.
TITLE Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice
JOURNAL Science 301 (5631), 376-379 (2003)
PUBMED 12869764
REFERENCE 2 (bases 1 to 1105)
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K., Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K., Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawagashira,N., Kawai,J., Kawamata,M., Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M., Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Kurosaki,T., Kusumegi,T., Li,C., Lu,M., Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A., Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakamura,M., Namiki,T., Narikawa,R., Niikura,J., Nishi,K., Nomura,K., Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M., Ooka,H., Osato,N., Ota,Y., Otomo,Y., Ryu,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Shibata,K., Shinagawa,A., Shiraki,T., Shishiki,T., Sogabe,Y., Sugano,S., Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,W., Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and Yoshimura,A.
TITLE Direct Submission
JOURNAL Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)
COMMENT This clone is one of the 28K full-length cDNA clones from japonica rice.
URL : http://cdna01.dna.affrc.go.jp/cDNA/
NIAS Rice Full-Length cDNA Project Team: Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T. and Yamamoto,M.
FAIS Genome Sequencing & Analysis Group: Otomo,Y., Iida,Y., Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M., Kobayashi,M., Kodama,T., Kurosaki,T., Kusumegi,T., Lu,M., Masuda,H., Miura,J., Mizuno,K., Narikawa,R., Niikura,J., Oka,M., Ryu,R., Sugano,S., Sugiyama,A., Suzuki,Y., Tsunoda,Y., Ueda,M., Xie,Q., Yokomizo,S., Yoshimura,A., Matsubara,K. and Murakami,K.
Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Iida,J., Imamura,K., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J., Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Osato,N., Ota,Y., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K., Yasunishi,A. and Hayashizaki,Y.
FEATURES Location/Qualifiers
source 1. .1105
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"

/clone="001-040-C08"
ORIGIN

Alignment Scores:
Pred. No.: 258 Length: 1105
Score: 42.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 15 Gaps: 0

US-09-858-332G-28 (1-6) x AK104803 (1-1105)

Qy 1 HisAsnHisAsnHisAsn 6
||||||||||||||||||
Db 322 CACAACCACAACCACAAC 339

RESULT 176
AY065142/c
LOCUS AY065142 1132 bp mRNA linear PLN 10-DEC-2001
DEFINITION Arabidopsis thaliana putative protein (At4g21810; T8O5.20) mRNA, complete cds.
ACCESSION AY065142
VERSION AY065142.1 GI:17473758
KEYWORDS FLI_CDNA.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 1132)
AUTHORS Nguyen,M., Karlin-Neumann,G., Southwick,A., Lam,B., Miranda,M., Palm,C.J., Bowser,L., Jones,T., Banh,J., Carninci,P., Chen,H., Cheuk,R., Chung,M.K., Hayashizaki,Y., Ishida,J., Kamiya,A., Kawai,J., Kim,C., Lin,J., Liu,S.X., Narusaka,M., Pham,P.K., Sakano,H., Sakurai,T., Satou,M., Seki,M., Shinn,P., Yamada,K., Shinozaki,K., Ecker,J., Theologis,A. and Davis,R.W.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-2001) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
COMMENT e-mail for correspondence: arab@sequence.stanford.edu

RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN Arabidopsis Full-Length cDNA'): Seki,M., Narusaka,M., Ishida,J., Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J., Hayashizaki,Y. and Shinozaki,K.

The Salk, Stanford, PGEC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Nguyen,M., Southwick,A., Karlin-Neumann,G., Lam,B., Miranda,M., Palm,C.J., Bowser,L., Jones,T., Banh,J., Chen,H., Cheuk,R., Chung,M.K., Kim,C., Lin,J., Liu,S.X., Pham,P.K., Sakano,H., Shinn,P., Yamada,K., Ecker,J., Theologis,A. and Davis,R.W.

Nguyen,M, (SSP/Stanford) and Seki,M. (RIKEN GSC) contributed equally to this work. Shinozaki,K. (RIKEN GSC) and Davis,R.W. (SSP/Stanford) contributed equally to this work as PIs.
FEATURES Location/Qualifiers
source 1. .1132
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="IV"
/clone="RAFL06-09-I21"
/ecotype="Columbia"
/note="This clone is in pBluescript"
gene 1. .1132
/gene="At4g21810; T8O5.20"
CDS 150. .884
/gene="At4g21810; T8O5.20"
/codon_start=1
/product="putative protein"
/protein_id="AAL38318.1"
/db_xref="GI:17473759"
/translation="MAQAVEEWYKQMPIITRSYLTAAVVTTVGCSLEIISPYNLYLNP
TLVVKQYQFWRLVTNFLYFRKMDLDFLFHMFFLARYCKLLEENSFRGKTADFLYMLLF
GATVLTGIVLIGGMIPYLSVSFSKIIFLSNSLTFMMVYVWSKQNPYIHMSFLGLFTFT
AAYLPWVLLGFSILVGASAWGDFLGMIAGHAYYFLAFVYPRMTDRRPLKTPSFLKALF
ADEPVVIARPEDVRFAHAPFDEIHQD"
ORIGIN

Alignment Scores:
Pred. No.: 264 Length: 1132
Score: 42.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 15 Gaps: 0

US-09-858-332G-28 (1-6) x AY065142 (1-1132)

Qy 1 HisAsnHisAsnHisAsn 6
||||||||||||||||||
Db 141 CACAATCACAATCACAAT 124

RESULT 177
AX164109
LOCUS AX164109 1154 bp DNA linear PAT 31-AUG-2001
DEFINITION Sequence 54 from Patent EP1118669.
ACCESSION AX164109
VERSION AX164109.1 GI:14545059
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1
AUTHORS Frenken,L.G., van der Logt,C.P., Jobling,S.A. and Teh,Y.M.
TITLE Production of camelid antibodies in plants
JOURNAL Patent: EP 1118669-A 54 25-JUL-2001;
UNILEVER PLC (GB) ; UNILEVER N.V. (NL)
FEATURES Location/Qualifiers
source 1. .1154
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="HCV33-hinge-CH2-CH3"
CDS 3. .1136
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC42712.1"
/db_xref="GI:14545060"
/translation="MEVQLQESGGGLVQAGGSLRLSCAASGRATSGHGHYGMGWFRQV
PGKEREFVAAIRWSGKETWYKDSVKGRFTISRDNAKTTVYLQMNSLKPEDTAVYYCAA
RPVRVDDISLPVGFDYWGQGTQVTVSSEPKTPKPQPQPQPQPQPNPTTESKCPKCPAP
ELLGGPSVFIFPPKPKDVLSISGRPEVTCVVVDVGQEDPEVSFNWYIDGAEVRTANTR
PKEEQFNSTYRVVSVLPIQHQDWLTGKEFKCKVNNKALPAPIEKTISKAKGQTREPQV
YALAPHREELAKDTVSVTCLVKGFYPPDINVEWQRNGQPESEGTYATTPPQLDNDGTY
FLYSKLSVGKNTWQRGETFTCVVMHEALHNHYTQKSITQSSGK"
ORIGIN

Alignment Scores:
Pred. No.: 269 Length: 1154
Score: 42.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-858-332G-28 (1-6) x AX164109 (1-1154)

Qy 1 HisAsnHisAsnHisAsn 6
||||||||||||||||||
Db 409 CACAACCACAACCACAAC 426

```
RESULT 178
AX164111
LOCUS       AX164111                1172 bp    DNA     linear   PAT 31-AUG-2001
DEFINITION  Sequence 56 from Patent EP1118669.
ACCESSION   AX164111
VERSION     AX164111.1  GI:14545061
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Frenken,L.G., van der Logt,C.P., Jobling,S.A. and Teh,Y.M.
  TITLE     Production of camelid antibodies in plants
  JOURNAL   Patent: EP 1118669-A 56 25-JUL-2001;
            UNILEVER PLC (GB) ; UNILEVER N.V. (NL)
FEATURES             Location/Qualifiers
     source          1. .1172
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="hinge-HCV33-CH2-CH3-SEKDEL"
     CDS             3. .1154
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAC42713.1"
                     /db_xref="GI:14545062"
                     /translation="MEVQLQESGGGLVQAGGSLRLSCAASGRATSGHGHYGMGWFRQV
                     PGKEREFVAAIRWSGKETWYKDSVKGRFTISRDNAKTTVYLQMNSLKPEDTAVYYCAA
                     RPVRVDDISLPVGFDYWGQGTQVTVSSEPKTPKPQPQPQPQPQPNPTTESKCPKCPAP
                     ELLGGPSVFIFPPKPKDVLSISGRPEVTCVVVDVGQEDPEVSFNWYIDGAEVRTANTR
                     PKEEQFNSTYRVVSVLPIQHQDWLTGKEFKCKVNNKALPAPIEKTISKAKGQTREPQV
                     YALAPHREELAKDTVSVTCLVKGFYPPDINVEWQRNGQPESEGTYATTPPQLDNDGTY
                     FLYSKLSVGKNTWQRGETFTCVVMHEALHNHYTQKSITQSSGKSEKDEL"
ORIGIN

Alignment Scores:
Pred. No.:              273             Length:         1172
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x AX164111 (1-1172)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        409 CACAACCACAACCACAAC 426

RESULT 179
AR550709
LOCUS       AR550709                1179 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 5840 from patent US 6747137.
ACCESSION   AR550709
VERSION     AR550709.1  GI:53943884
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1179)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 5840 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .1179
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              274             Length:         1179
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x AR550709 (1-1179)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        301 CATAACCATAACCATAAC 318

RESULT 180
CQ803932
LOCUS       CQ803932                1206 bp    DNA     linear   PAT 10-MAY-2004
DEFINITION  Sequence 343 from Patent WO2004035798.
ACCESSION   CQ803932
VERSION     CQ803932.1  GI:47110600
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Inze,D., de Veylder,L. and Vlieghe,K.
  TITLE     Identification of novel e2f target genes and use thereof
  JOURNAL   Patent: WO 2004035798-A 343 29-APR-2004;
            CropDesign N.V. (BE)
FEATURES             Location/Qualifiers
     source          1. .1206
                     /organism="Arabidopsis thaliana"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:3702"
ORIGIN

Alignment Scores:
Pred. No.:              280             Length:         1206
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x CQ803932 (1-1206)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        880 CACAATCATAATCACAAT 897

RESULT 181
AK068117
LOCUS       AK068117                1233 bp    mRNA    linear   PLN 24-JUL-2003
DEFINITION  Oryza sativa (japonica cultivar-group) cDNA clone:J013134A06, full
            insert sequence.
ACCESSION   AK068117
VERSION     AK068117.1  GI:32978135
KEYWORDS    FLI_CDNA; CAP trapper.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   The Rice Full-Length cDNA Consortium, National Institute of
            Agrobiological Sciences Rice Full-Length cDNA Project Team:,
            Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K.,
            Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I.,
            Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C.,
            Ohtsuki,K., Shishiki,T., Foundation of Advancement of International
            Science Genome Sequencing & Analysis Group:, Otomo,Y., Murakami,K.,
```

Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y., Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M., Narikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Niikura,J., Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Miura,J., Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN:, Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Osato,N., Ota,Y., Saito,R., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Yoshino,M. and Hayashizaki,Y.

TITLE Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice

JOURNAL Science 301 (5631), 376-379 (2003)

PUBMED 12869764

REFERENCE 2 (bases 1 to 1233)

AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K., Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K., Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawagashira,N., Kawai,J., Kawamata,M., Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M., Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Kurosaki,T., Kusumegi,T., Li,C., Lu,M., Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A., Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakamura,M., Namiki,T., Narikawa,R., Niikura,J., Nishi,K., Nomura,K., Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M., Ooka,H., Osato,N., Ota,Y., Otomo,Y., Ryu,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Shibata,K., Shinagawa,A., Shiraki,T., Shishiki,T., Sogabe,Y., Sugano,S., Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,W., Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and Yoshimura,A.

TITLE Direct Submission

JOURNAL Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)

COMMENT This clone is one of the 28K full-length cDNA clones from japonica rice.
URL : http://cdna01.dna.affrc.go.jp/cDNA/
NIAS Rice Full-Length cDNA Project Team: Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T. and Yamamoto,M.
FAIS Genome Sequencing & Analysis Group: Otomo,Y., Iida,Y., Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M., Kobayashi,M., Kodama,T., Kurosaki,T., Kusumegi,T., Lu,M., Masuda,H., Miura,J., Mizuno,K., Narikawa,R., Niikura,J., Oka,M., Ryu,R., Sugano,S., Sugiyama,A., Suzuki,Y., Tsunoda,Y., Ueda,M., Xie,Q., Yokomizo,S., Yoshimura,A., Matsubara,K. and Murakami,K.
Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Iida,J., Imamura,K., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J., Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Osato,N., Ota,Y., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K., Yasunishi,A. and Hayashizaki,Y.

FEATURES Location/Qualifiers

source 1. .1233
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="J013134A06"

ORIGIN

Alignment Scores:

Pred. No.:	286	Length:	1233
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	15	Gaps:	0

US-09-858-332G-28 (1-6) x AK068117 (1-1233)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        448 CACAACCACAACCACAAC 465
```

RESULT 182
AY091398
LOCUS AY091398 1237 bp mRNA linear PLN 18-SEP-2002
DEFINITION Arabidopsis thaliana unknown protein (At1g78420) mRNA, complete cds.
ACCESSION AY091398
VERSION AY091398.1 GI:20259242
KEYWORDS FLI_CDNA.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 1237)

AUTHORS Yamada,K., Banh,J., Chan,M.M., Chang,C.H., Chang,E., Dale,J.M., Deng,J.M., Goldsmith,A.D., Lee,J.M., Onodera,C.S., Quach,H.L., Tang,C., Toriumi,M., Wu,H.C., Yamamura,Y., Yu,G., Bowser,L., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B., Lin,J., Meyers,M.C., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.

TITLE Arabidopsis Open Reading Frame (ORF) Clones

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1237)

AUTHORS Yamada,K., Banh,J., Chan,M.M., Chang,C.H., Chang,E., Dale,J.M., Deng,J.M., Goldsmith,A.D., Lee,J.M., Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wu,H.C., Yamamura,Y., Yu,G., Bowser,L., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B., Lin,J., Meyers,M.C., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.

TITLE Direct Submission

JOURNAL Submitted (20-MAR-2002) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA

COMMENT The RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN Arabidopsis Full-Length cDNA'): Seki,M., Narusaka,M., Ishida,J., Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai, J., Hayashizaki,Y. and Shinozaki,K.

The Salk, Stanford, PGEC (SSP) Consortium members constructed and sequenced the pUNI (ORF) clones using the RAFL cDNAs: Yamada,K., Banh,J., Chan,M.M., Chang,C.H., Chang,E., Dale,J.M., Deng,J.M., Goldsmith,A.D., Lee,J.M., Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wu,H.C., Yamamura,Y., Yu,G., Bowser,L., Chen,H., Cheuk,R., Jones,T., Karlin-Neumann,G., Kim,C., Lam,B., Lin,J., Meyers,M.C., Miranda,M., Nguyen,M., Palm,C.J., Shinn,P., Southwick,A., Davis,R.W., Ecker,J.R. and Theologis,A.

Yamada,K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP

```
           /PGEC) contributed equally to this work as PIs.

           Annotation is based on the January 2002 version of the Arabidopsis
           genome submitted to GenBank.
FEATURES             Location/Qualifiers
     source          1. .1237
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /chromosome="1"
                     /clone="U10670"
                     /ecotype="Columbia"
                     /note="This clone is in pUNI 51."
     gene            1. .1237
                     /gene="At1g78420"
     CDS             1. .1206
                     /gene="At1g78420"
                     /codon_start=1
                     /evidence=experimental
                     /product="unknown protein"
                     /protein_id="AAM14337.1"
                     /db_xref="GI:20259243"
                     /translation="MGNKLGRKRQVVEERYTKPQGLYVNKDVDVKKLRKLIVESKLAP
                     CYPGDDESCHDLEECPICFLYYPSLNRSRCCMKSICTECFLQMKNPNSARPTQCPFCK
                     TPNYAVEYRGVKSKEEKGIEQVEEQRVIEAKIRMRQKEMQDDEEKMQKRLESCSSSTS
                     AMTGEMEYGSTSAISYNSLMDDGEIAPSQNASVVRQHSRPRGNREDEVDVDLEELMVM
                     EAIWLSVQETGTQRNSASGEITSSRQYVTDNHSYVSSPPRVTPIVEPATPSSSSGGLS
                     CAISALAERQMVGESSSHNHNHNVNVSSYSMLPGNCDSYYDIEQEVDGIDNHHHHRHH
                     YEMGETGSSNSYVSSYMTGEGFHNFPPPPPLVIVPESFEEQMMMAMAVSMAEVHATTT
                     CAPTEVTWQ"
     3'UTR           1207. .1237
                     /gene="At1g78420"
ORIGIN

Alignment Scores:
Pred. No.:              287             Length:         1237
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x AY091398 (1-1237)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        880 CACAATCATAATCACAAT 897

RESULT 183
DQ120961
LOCUS       DQ120961               1239 bp    DNA     linear   VRT 02-AUG-2005
DEFINITION  Hucho taimen clone ZB-12 microsatellite sequence.
ACCESSION   DQ120961
VERSION     DQ120961.1  GI:71385014
KEYWORDS    .
SOURCE      Hucho taimen
  ORGANISM  Hucho taimen
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Protacanthopterygii; Salmoniformes; Salmonidae; Hucho.
REFERENCE   1  (bases 1 to 1239)
  AUTHORS   Sun,X., Tong,G., Lu,C., Kuang,Y. and Liang,L.
  TITLE     Microsatellite Enrichment by Magnetic Beads in taimen
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1239)
  AUTHORS   Sun,X., Tong,G., Lu,C., Kuang,Y. and Liang,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUL-2005) Biotech Breeding Lab, Heilongjiang River
            Fishery Research Institute of Chinese Academy of Fishery Science,
            No. 43 Songfa Road Daoli District, Harbin, Heilongjiang 150070,
            China
FEATURES             Location/Qualifiers
     source          1. .1239
                     /organism="Hucho taimen"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:201498"
                     /clone="ZB-12"
     repeat_region   224. .747
                     /note="microsatellite"
                     /rpt_type=tandem
ORIGIN

Alignment Scores:
Pred. No.:              288             Length:         1239
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     5               Gaps:           0

US-09-858-332G-28 (1-6) x DQ120961 (1-1239)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        654 CATAACCACAACCACAAC 671

RESULT 184
AF079247/c
LOCUS       AF079247               1258 bp    DNA     linear   PLN 06-JAN-1999
DEFINITION  Cymbopogon commutatus granule-bound starch synthase (waxy) gene,
            partial cds.
ACCESSION   AF079247
VERSION     AF079247.1  GI:3493018
KEYWORDS    .
SOURCE      Cymbopogon commutatus
  ORGANISM  Cymbopogon commutatus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Cymbopogon.
REFERENCE   1  (bases 1 to 1258)
  AUTHORS   Mason-Gamer,R.J., Weil,C.F. and Kellogg,E.A.
  TITLE     Granule-bound starch synthase: structure, function, and
            phylogenetic utility
  JOURNAL   Mol. Biol. Evol. 15 (12), 1658-1673 (1998)
   PUBMED   9866201
REFERENCE   2  (bases 1 to 1258)
  AUTHORS   Mason-Gamer,R.J., Weil,C.F. and Kellogg,E.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-1998) Biological Sciences, University of Idaho,
            Life Sciences South, Moscow, ID 83844-3051, USA
FEATURES             Location/Qualifiers
     source          1. .1258
                     /organism="Cymbopogon commutatus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:79834"
     gene            <1. .>1258
                     /gene="waxy"
     mRNA            join(<1. .120,203. .379,484. .675,813. .899,988. .1116,
                     1200. .>1258)
                     /gene="waxy"
                     /product="granule-bound starch synthase"
     CDS             join(<1. .120,203. .379,484. .675,813. .899,988. .1116,
                     1200. .>1258)
                     /gene="waxy"
                     /note="GBSSI"
                     /codon_start=1
                     /product="granule-bound starch synthase"
                     /protein_id="AAD02967.1"
                     /db_xref="GI:3493019"
                     /translation="CELDNIMRLTGITGIVNGMDVSEWDPSKDKYIAVKYDVSTAMEA
                     KALNKEALQAEVGLPVDRKVPLVAFIGRLEEQKGPDVMAAAIPQLMEEDVQIVLLGTG
                     KKKFERMLMSLEEKYPDKVRAVVKFNAALAHHIMAGADVLAVTSRFEPCGLIQLQGMR
                     YGTPCACASTGGLVDTIIEGKTGFHMGRLSVDCNVVEPADVKKVATTLKRAIKVVGTP
                     AYEEMVKNCMIQDLSWKGPAKNWENVLLSLGVAGGEP"
ORIGIN
```

```
Alignment Scores:
Pred. No.:                292             Length:          1258
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       15              Gaps:            0

US-09-858-332G-28 (1-6) x AF079247 (1-1258)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        789 CACAATCACAATCACAAC 772

RESULT 185
AF462841
LOCUS       AF462841               1258 bp    mRNA    linear   PLN 08-JAN-2002
DEFINITION  Arabidopsis thaliana At3g12480/MQC3.32 mRNA, complete cds.
ACCESSION   AF462841
VERSION     AF462841.1  GI:18087596
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1258)
  AUTHORS   Cheuk,R., Chen,H., Kim,C.J., Meyers,M.C., Banh,J., Bowser,L.,
            Carninci,P., Chang,E., Dale,J.M., Goldsmith,A.D., Hayashizaki,Y.,
            Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J.,
            Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M.,
            Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M.,
            Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C., Yamada,K.,
            Yamamura,Y., Yu,G., Yu,S., Shinozaki,K., Davis,R.W., Theologis,A.
            and Ecker,J.R.
  TITLE     Arabidopsis cDNA clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1258)
  AUTHORS   Cheuk,R., Chen,H., Kim,C.J., Meyers,M.C., Banh,J., Bowser,L.,
            Carninci,P., Chang,E., Dale,J.M., Goldsmith,A.D., Hayashizaki,Y.,
            Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J.,
            Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M.,
            Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M.,
            Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C., Yamada,K.,
            Yamamura,Y., Yu,G., Yu,S., Shinozaki,K., Davis,R.W., Theologis,A.
            and Ecker,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-DEC-2001) Salk Institute Genomic Analysis Laboratory
            (SIGnAL), Plant Biology Laboratory, The Salk Institute for
            Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
            USA
COMMENT     RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis Full-Length cDNA') :  Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P.,  Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members carried out the
            sequencing and annotation of the RAFL cDNAs: Cheuk,R., Chen,H.,
            Kim,C.J., Meyers,M.C., Shinn,P., Banh,J. Bowser,L., Dale,J.M.,
            Goldsmith,A.D., Jiang,P.X., Jones,T., Karlin-Neumann,G., Lam,B.,
            Lee,J.M., Lin,J., Liu,S.X., Miranda,M., Nguyen,M., Onodera,C.S.,
            Palm,C.J., Pham,P.K., Quach,H.L., Southwick,A., Tang,C.C.,
            Toriumi,M., Yamada,K., Yamamura, Y., Yu,G., Yu,S., Davis,R.W.,
            Theologis,A., and Ecker,J.R.

            Cheuk,R. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to
            this work.  Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk)
            contributed equally to this work as PIs.
FEATURES              Location/Qualifiers
     source           1. .1258
                      /organism="Arabidopsis thaliana"
                      /mol_type="mRNA"
                      /db_xref="taxon:3702"
                      /chromosome="3"
                      /clone="RAFL05-14-K14(R21457)"
                      /ecotype="Columbia"
     5'UTR            1. .228
     CDS              229. .1110
                      /note="unknown protein"
                      /codon_start=1
                      /product="At3g12480/MQC3.32"
                      /protein_id="AAL58929.1"
                      /db_xref="GI:18087597"
                      /translation="MRKKLDTRFPAARIKKIMQADEDVGKIALAVPVLVSKSLELFLQ
                      DLCDRTYEITLERGAKTVSSLHLKHCVERYNVFDFLREVVSKVPDYGHSQGQGHGDVT
                      MDDRSISKRRKPISDEVNDSDEEYKKSKTQEIGSAKTSGRGGRGRGRGRGGRAAKA
                      AEREGLNREMEVEAANSGQPPPEDNVKMHASESSPQEDEKKGIDGTAASNEDTKQHLQ
                      SPKEGIDFDLNAESLDLNETKLAPATGTTTTTAATDSEEYSGWPMMDISKMDPAQLA
                      SLGKRIDEDEEDYDEEG"
     3'UTR            1111. .1258
ORIGIN

Alignment Scores:
Pred. No.:                292             Length:          1258
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       15              Gaps:            0

US-09-858-332G-28 (1-6) x AF462841 (1-1258)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        963 CACAACCACAACCACAAC 980

RESULT 186
AR550357
LOCUS       AR550357               1263 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 5488 from patent US 6747137.
ACCESSION   AR550357
VERSION     AR550357.1  GI:53943532
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1263)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 5488 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES              Location/Qualifiers
     source           1. .1263
                      /organism="unknown"
                      /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:                293             Length:          1263
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       6               Gaps:            0

US-09-858-332G-28 (1-6) x AR550357 (1-1263)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        383 CACAACCACAACCACAAC 400

RESULT 187
ZFVICGN/c
LOCUS       ZFVICGN                1304 bp    DNA     linear   PLN 18-APR-2005
```

```
DEFINITION  Z.furfuracea gene for vicilin.
ACCESSION   Z54365
VERSION     Z54365.1  GI:1019832
KEYWORDS    7S globulin; vicilin.
SOURCE      Zamia furfuracea
  ORGANISM  Zamia furfuracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Cycadophyta; Cycadales; Zamiaceae; Zamia.
REFERENCE   1  (bases 1 to 1304)
  AUTHORS   Braun,H.
  JOURNAL   Thesis (1995) Molekulare Genetik, Institut fuer Pflanzengenetik und
            Kulturpflanzenforschung
REFERENCE   2  (bases 1 to 1304)
  AUTHORS   Braun,H., Czihal,A., Shutov,A.D. and Baumlein,H.
  TITLE     A vicilin-like seed protein of cycads: similarity to
            sucrose-binding proteins
  JOURNAL   Plant Mol. Biol. 31 (1), 35-44 (1996)
   PUBMED   8704157
REFERENCE   3  (bases 1 to 1304)
  AUTHORS   Braun,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-1995) Braun H., Institut fuer Pflanzengenetik und
            Kulturpflanzenforschung, Molekulare Genetik, Correnssstr.3,
            Gatersleben, Germany, D-06466
COMMENT     Accession number incorrectly printed in journal.
FEATURES             Location/Qualifiers
     source          1. .1304
                     /organism="Zamia furfuracea"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:42329"
                     /clone="gZ1,3/10"
                     /tissue_type="leaf"
                     /clone_lib="PCR"
                     /dev_stage="middle"
     CDS             join(<1. .29,112. .290,366. .446,539. .802,912. .1191,
                     1277. .>1304)
                     /function="seed storage protein"
                     /standard_name="7s globulin"
                     /note="partial genomic PCR-fragment. Corresponding cDNA:
                     Z50791"
                     /codon_start=2
                     /product="vicilin"
                     /protein_id="CAA91188.1"
                     /db_xref="GI:1019833"
                     /db_xref="GOA:Q41726"
                     /db_xref="InterPro:IPR006045"
                     /db_xref="InterPro:IPR007113"
                     /db_xref="UniProt/TrEMBL:Q41726"
                     /translation="ADLALYVTGGRGRVAFVHEERLVERQLRDGDVYAIAAGIPFYIL
                     NTDDSRRLFIHCLLRTQCSTTGLYESFYVVGGRNPQNVLSQFSEDVLQAAFNSSKAVL
                     DPMLVSGFNRGAIITVSREQMERLSRGRIKGFGGSEEPQPFNLLYRNPDFSNNNGEIF
                     TADAADHRVLRRLNVGVQLLNLKPRSMTAPHYDTRSTRIGIVRNGRGILELVRPQEQE
                     QQQQQQGPTYQKLRANLNPGTVFLTRPGYPSTVIASGNEALQILYFDINSQGNRRQFL
                     AGRSNVLRYLP"
     exon            1. .29
     intron          30. .111
     exon            112. .290
     intron          291. .365
     exon            366. .446
     intron          447. .538
     exon            539. .802
     intron          803. .911
     exon            912. .1191
     intron          1192. .1276
ORIGIN

Alignment Scores:
Pred. No.:              302           Length:         1304
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     15            Gaps:           0
```

```
US-09-858-332G-28 (1-6) x ZFVICGN (1-1304)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         94 CACAATCACAATCACAAT 77

RESULT 188
AR554685
LOCUS       AR554685               1317 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 9816 from patent US 6747137.
ACCESSION   AR554685
VERSION     AR554685.1  GI:53947860
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1317)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 9816 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .1317
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              305           Length:         1317
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     6             Gaps:           0

US-09-858-332G-28 (1-6) x AR554685 (1-1317)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        298 CATAATCATAATCATAAT 315

RESULT 189
BT014236
LOCUS       BT014236               1327 bp    mRNA    linear   PLN 11-MAY-2004
DEFINITION  Lycopersicon esculentum clone 133435F, mRNA sequence.
ACCESSION   BT014236
VERSION     BT014236.1  GI:47105651
KEYWORDS    FLI_CDNA.
SOURCE      Lycopersicon esculentum (Solanum lycopersicum)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 1327)
  AUTHORS   Kirkness,E.F., Wang,W. and Vazeille,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAY-2004) The Institute for Genomic Research, 9712
            Medical Center Drive, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .1327
                     /organism="Lycopersicon esculentum"
                     /mol_type="mRNA"
                     /db_xref="taxon:4081"
                     /clone="133435F"
                     /tissue_type="fruit"
                     /note="TZXHK76"
ORIGIN

Alignment Scores:
Pred. No.:              307           Length:         1327
Score:                  42.00         Matches:        6
```

```
Percent Similarity:     100.00%         Conservative:  0
Best Local Similarity:  100.00%         Mismatches:    0
Query Match:            100.00%         Indels:        0
DB:                     15              Gaps:          0

US-09-858-332G-28 (1-6) x BT014236 (1-1327)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        253 CACAACCACAACCACAAC 270

RESULT 190
AR549628
LOCUS       AR549628               1350 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 4759 from patent US 6747137.
ACCESSION   AR549628
VERSION     AR549628.1  GI:53942803
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1350)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 4759 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .1350
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              312             Length:        1350
Score:                  42.00           Matches:       6
Percent Similarity:     100.00%         Conservative:  0
Best Local Similarity:  100.00%         Mismatches:    0
Query Match:            100.00%         Indels:        0
DB:                     6               Gaps:          0

US-09-858-332G-28 (1-6) x AR549628 (1-1350)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        998 CACAACCACAATCACAAT 1015

RESULT 191
AF139533/c
LOCUS       AF139533               1377 bp    mRNA    linear   PLN 13-AUG-1999
DEFINITION  Liquidambar styraciflua caffeate O-methyltransferase (COMT) mRNA,
            complete cds.
ACCESSION   AF139533
VERSION     AF139533.1  GI:5731999
KEYWORDS    .
SOURCE      Liquidambar styraciflua (sweet gum)
  ORGANISM  Liquidambar styraciflua
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            Saxifragales; Hamamelidaceae; Liquidambar.
REFERENCE   1  (bases 1 to 1377)
  AUTHORS   Osakabe,K., Tsao,C.C., Li,L., Popko,J.L., Umezawa,T.,
            Carraway,D.T., Smeltzer,R.H., Joshi,C.P. and Chiang,V.L.
  TITLE     Coniferyl aldehyde 5-hydroxylation and methylation direct syringyl
            lignin biosynthesis in angiosperms
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (16), 8955-8960 (1999)
   PUBMED   10430877
REFERENCE   2  (bases 1 to 1377)
  AUTHORS   Osakabe,K., Tsao,C.C., Li,L., Popko,J.L., Umezawa,T.,
            Carraway,D.T., Smeltzer,R.H., Joshi,C.P. and Chiang,V.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-1999) Institute of Wood Research, Michigan
            Technological University, 1400 Townsend Drive, Houghton, MI 49931,
            USA
FEATURES             Location/Qualifiers
     source          1. .1377
                     /organism="Liquidambar styraciflua"
                     /mol_type="mRNA"
                     /db_xref="taxon:4400"
                     /tissue_type="developing xylem"
     gene            1. .1377
                     /gene="COMT"
     CDS             67. .1170
                     /gene="COMT"
                     /EC_number="2.1.1.6"
                     /codon_start=1
                     /product="caffeate O-methyltransferase"
                     /protein_id="AAD48913.1"
                     /db_xref="GI:5732000"
                     /translation="MGSTSETKMSPSEAAAAAEEEAFVFAMQLTSASVLPMVLKSAIEL
                     DVLEIMAKAGPGAHISTSDIASKLPTKNPDAAVMLDRMLRLLASYSVLTCSLRTLPDG
                     KIERLYGLAPVCKFLTRNDDGVSIAALSLMNQDKVLMESWYHLTEAVLEGGIPFNKAY
                     GMTAFEYHGTDPRFNTVFNNGMSNHSTITMKKILETYKGFEGLGSVVDVGGGTGAHLN
                     MIIAKYPMIKGINFDLPHVIEEAPSYPGVEHVGGDMFVSVPKGDAIFMKWICHDWSDE
                     HCLKFLKKCYEALPTNGKVILAECILPVAPDASLPTKAVVHIDVIMLAHNPGGKERTE
                     KEFEALAKGAGFEGFRVALCAYNTWIIEFLKKI"
ORIGIN

Alignment Scores:
Pred. No.:              318             Length:        1377
Score:                  42.00           Matches:       6
Percent Similarity:     100.00%         Conservative:  0
Best Local Similarity:  100.00%         Mismatches:    0
Query Match:            100.00%         Indels:        0
DB:                     15              Gaps:          0

US-09-858-332G-28 (1-6) x AF139533 (1-1377)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1238 CACAATCACAATCACAAT 1221

RESULT 192
BT002883/c
LOCUS       BT002883               1394 bp    mRNA    linear   PLN 15-JAN-2003
DEFINITION  Arabidopsis thaliana clone RAFL15-04-C12 (R20378) putative copper
            chaperone (CCH) protein (At1g56210) mRNA, complete cds.
ACCESSION   BT002883
VERSION     BT002883.1  GI:27754504
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1394)
  AUTHORS   Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,
            Onodera,C.S., Quach,H.L., Tang,C., Toriumi,M., Wong,C., Wu,H.C.,
            Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,
            Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M.,
            Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P.,
            Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W.,
            Ecker,J.R. and Theologis,A.
  TITLE     Arabidopsis Full Length cDNA Clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1394)
  AUTHORS   Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,
            Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C.,
            Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,
            Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M.,
            Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P.,
            Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W.,
            Ecker,J.R. and Theologis,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JAN-2003) Plant Gene Expression Center, 800 Buchanan
            Street, Albany, CA 94710, USA
```

```
COMMENT     RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis Full-Length cDNA'): Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members carried out the
            sequencing and annotation of the RAFL cDNAs: Yamada,K., Chan,M.M.,
            Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Onodera,C.S.,
            Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C., Yu,G.,
            Yuan,S., Chen,H., Cheuk,R., Jones,T., Kim,C.J., Nguyen,M.,
            Palm,C.J., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Davis,R.W.,
            Ecker,J.R. and Theologis,A.

            Yamada,K. (SSP/PGEC) and Seki,M. (RIKEN GSC) contributed equally to
            this work. Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP/PGEC)
            contributed equally to this work as PIs.

            Annotation based on July 2002 version of the Arabidopsis genome
            submitted to Genbank.
FEATURES              Location/Qualifiers
     source           1. .1394
                      /organism="Arabidopsis thaliana"
                      /mol_type="mRNA"
                      /db_xref="taxon:3702"
                      /chromosome="1"
                      /clone="RAFL15-04-C12 (R20378)"
                      /ecotype="Columbia"
                      /note="This clone is in a modified pBluescript vector2
                      (lambda PS) as a BamHI/XhoI insert."
     gene             1. .1394
                      /gene="At1g56210"
     5'UTR            1. .89
                      /gene="At1g56210"
     CDS              90. .1184
                      /gene="At1g56210"
                      /codon_start=1
                      /evidence=experimental
                      /product="putative copper chaperone (CCH) protein"
                      /protein_id="AAO22700.1"
                      /db_xref="GI:27754505"
                      /translation="MATDEMKSETKKTEHKQKQSTQIKQDLPPPTIPPLPLPYKSCTL
                      KVSIHCEGCKRKVKKILTSIEGVFKVDIDVKQHKVTVIGIISPEILLKKLNKAGKNAE
                      QLPEIPDPVDNKPKPVDPKEKNKKKKKEEKVQITNEATSSGIDNPEKPGSGECDKPES
                      EKPVDEKCLSGDGGETSGPVKEEKKDVLKEKDSGKEESPSPPADSSAPAAEKKAEDTG
                      GAVPDNGKVGKKKKKKGQSLATTNNPTDGPARTQSLPPPTATDYDRPINQINDHNITT
                      TNNPPRHDLYPYPAQGYYAPQVVYGVSYNVAQPPVSVDAASYYTPPPPYSYAYMHNGY
                      QPSDQNPCQPRPSDSFELFSDENPNGCAIM"
     misc_difference  906
                      /gene="At1g56210"
                      /note="compared to genomic sequence resulting in an amino
                      acid sequence difference"
                      /replace="c"
     misc_difference  984
                      /gene="At1g56210"
                      /note="compared to genomic sequence resulting in an amino
                      acid sequence difference"
                      /replace="a"
     3'UTR            1185. .1394
                      /gene="At1g56210"
     misc_difference  1333. .1334
                      /gene="At1g56210"
                      /note="compared to genomic sequence"
                      /replace="aat"
ORIGIN

Alignment Scores:
Pred. No.:              322          Length:         1394
Score:                  42.00        Matches:        6
Percent Similarity:     100.00%      Conservative:   0
Best Local Similarity:  100.00%      Mismatches:     0
Query Match:            100.00%      Indels:         0
DB:                     15           Gaps:           0

US-09-858-332G-28 (1-6) x BT002883 (1-1394)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1350 CATAATCATAATCATAAT 1333

RESULT 193
AK108026
LOCUS       AK108026               1418 bp    mRNA    linear   PLN 24-JUL-2003
DEFINITION  Oryza sativa (japonica cultivar-group) cDNA clone:002-137-A02, full
            insert sequence.
ACCESSION   AK108026
VERSION     AK108026.1  GI:32993235
KEYWORDS    FLI_CDNA; oligo capping.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   The Rice Full-Length cDNA Consortium, National Institute of
            Agrobiological Sciences Rice Full-Length cDNA Project Team:,
            Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K.,
            Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I.,
            Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C.,
            Ohtsuki,K., Shishiki,T., Foundation of Advancement of International
            Science Genome Sequencing & Analysis Group:, Otomo,Y., Murakami,K.,
            Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y.,
            Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M.,
            Narikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Niikura,J.,
            Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Miura,J.,
            Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN:,
            Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuda,S.,
            Hara,A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M.,
            Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Osato,N., Ota,Y.,
            Saito,R., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
            Yoshino,M. and Hayashizaki,Y.
  TITLE     Collection, mapping, and annotation of over 28,000 cDNA clones from
            japonica rice
  JOURNAL   Science 301 (5631), 376-379 (2003)
   PUBMED   12869764
REFERENCE   2  (bases 1 to 1418)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K.,
            Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiraoka,T.,
            Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K.,
            Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I.,
            Kanagawa,S., Katoh,H., Kawagashira,N., Kawai,J., Kawamata,M.,
            Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M.,
            Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Kurosaki,T., Kusumegi,T., Li,C., Lu,M.,
            Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A.,
            Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakamura,M.,
            Namiki,T., Narikawa,R., Niikura,J., Nishi,K., Nomura,K.,
            Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M., Ooka,H.,
            Osato,N., Ota,Y., Otomo,Y., Ryu,R., Saitoh,H., Sakai,C., Sakai,K.,
            Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Shibata,K.,
            Shinagawa,A., Shiraki,T., Shishiki,T., Sogabe,Y., Sugano,S.,
            Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y.,
            Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A.,
            Toya,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,W.,
            Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and
            Yoshimura,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of
            Agrobiological Sciences, Department of Molecular Genetics, Head of
            Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki
            305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp,
            Tel:81-29-838-7007, Fax:81-29-838-7007)
COMMENT     This clone is one of the 28K full-length cDNA clones from japonica
            rice.
            URL : http://cdna01.dna.affrc.go.jp/cDNA/
```

NIAS Rice Full-Length cDNA Project Team: Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T. and Yamamoto,M.
FAIS Genome Sequencing & Analysis Group: Otomo,Y., Iida,Y., Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M., Kobayashi,M., Kodama,T., Kurosaki,T., Kusumegi,T., Lu,M., Masuda,H., Miura,J., Mizuno,K., Narikawa,R., Niikura,J., Oka,M., Ryu,R., Sugano,S., Sugiyama,A., Suzuki,Y., Tsunoda,Y., Ueda,M., Xie,Q., Yokomizo,S., Yoshimura,A., Matsubara,K. and Murakami,K.
Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Iida,J., Imamura,K., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J., Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Osato,N., Ota,Y., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K., Yasunishi,A. and Hayashizaki,Y.

FEATURES Location/Qualifiers
source 1. .1418
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="002-137-A02"

ORIGIN

Alignment Scores:

Pred. No.:	327	Length:	1418
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	15	Gaps:	0

US-09-858-332G-28 (1-6) x AK108026 (1-1418)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        564 CACAACCACAACCACAAC 581
```

RESULT 194
AR549760
LOCUS AR549760 1431 bp DNA linear PAT 08-OCT-2004
DEFINITION Sequence 4891 from patent US 6747137.
ACCESSION AR549760
VERSION AR549760.1 GI:53942935
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1431)
AUTHORS Weinstock,K.G. and Bush,D.
TITLE Nucleic acid sequences relating to Candida albicans for diagnostics and therapeutics
JOURNAL Patent: US 6747137-A 4891 08-JUN-2004;
Genome Therapeutics Corporation; Waltham, MA

FEATURES Location/Qualifiers
source 1. .1431
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Alignment Scores:

Pred. No.:	330	Length:	1431
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-09-858-332G-28 (1-6) x AR549760 (1-1431)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        816 CACAACCACAACCACAAC 833
```

RESULT 195
PUFFC4/c
LOCUS PUFFC4 1455 bp mRNA linear INV 19-DEC-1995
DEFINITION Bradysia hygida PC4-1 mRNA, complete cds.
ACCESSION U13883
VERSION U13883.1 GI:540082
KEYWORDS .
SOURCE Bradysia hygida
ORGANISM Bradysia hygida
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Sciaroidea; Sciaridae; Bradysia.
REFERENCE 1 (bases 1 to 1455)
AUTHORS Monesi,N., Fernandez,M.A., Fontes,A.M., Basso,L.R. Jr., Nakanishi,Y., Baron,B., Buttin,G. and Paco-Larson,M.L.
TITLE Molecular characterization of an 18 kb segment of DNA puff C4 of Bradysia hygida (Diptera, sciaridae)
JOURNAL Chromosoma 103 (10), 715-724 (1995)
PUBMED 7664619
REFERENCE 2 (bases 1 to 1455)
AUTHORS Paco-Larson,M.L.
TITLE Direct Submission
JOURNAL Submitted (22-AUG-1994) Maria Luisa M. Paco-Larson, Morfologia, Faculdade de Medicina de Ribeirao Preto, Universidade de Sao Paulo, Av. Bandeirantes 3900, Ribeirao Preto, Sao Paulo, Brasil, 14049-900

FEATURES Location/Qualifiers
source 1. .1455
/organism="Bradysia hygida"
/mol_type="mRNA"
/db_xref="taxon:35572"
/chromosome="C"
/map="region C4"
/clone="pcBhC4"
/tissue_type="salivary gland"
/dev_stage="instar, last larval"
gene 1. .1455
/gene="PC4-1"
CDS 22. .1374
/gene="PC4-1"
/codon_start=1
/protein_id="AAA83554.1"
/db_xref="GI:540083"
/translation="MKCFLILVIAVVALQCVSANPSRKRDCDCDCDCDKRRRIDDDCD
DYRFHFDDDDCFDNNKIEFRGHDDCDDVFDFGHKHGHGHGHGHGHGHGHGNNHHGGRKHE
HGNRHQHRGQKLHGEFSKIEKHGSSFGGVAGSLAQIISTISCLTGSATQILSQIQIES
LIKVVIEVVNVTSCVLGSATQLVVSAIESVLESLTCLSRIIEKITEVAIEVIEKAIEQ
VVCAISALIDTVAVIAEAVVDIAVEAVAEALSTLFLVLNVVSYAVETVICLVVDIVEE
VFQIVSETIEVVIGAIFTAVEALFEVVQKISCLFEVAIGVVVSLIGVVSQITQVLSSV
ICTLTDTLSVVINIVTGAVGSVCGLVGSVAKITSVVTSTVGNVVKGLTGALSGLTGGL
NLGGGLRLGGGKGGSVKQNIGGKSGFDGGFSGGHVMTSAFAQGKNVGFGKGHGGFNLF
"
polyA_signal 1433. .1438
/gene="PC4-1"
polyA_site 1455
/gene="PC4-1"

ORIGIN

Alignment Scores:

Pred. No.:	336	Length:	1455
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0

```
Query Match:            100.00%       Indels:        0
DB:                     2             Gaps:          0

US-09-858-332G-28 (1-6) x PUFFC4 (1-1455)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        115 CACAATCACAATCACAAT 98

RESULT 196
AX489036
LOCUS       AX489036               1476 bp    DNA     linear   PAT 16-AUG-2002
DEFINITION  Sequence 6336 from Patent WO02053728.
ACCESSION   AX489036
VERSION     AX489036.1  GI:22323048
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1
  AUTHORS   Roemer,T., Jiang,B., Boone,C., Bussey,H. and Ohlsen,K.L.
  TITLE     Gene disruption methodologies for drug target discovery
  JOURNAL   Patent: WO 02053728-A 6336 11-JUL-2002;
            Elitra Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .1476
                     /organism="Candida albicans"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:5476"
ORIGIN

Alignment Scores:
Pred. No.:              340           Length:        1476
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-28 (1-6) x AX489036 (1-1476)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        782 CACAATCACAATCACAAT 799

RESULT 197
AR549490
LOCUS       AR549490               1479 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 4621 from patent US 6747137.
ACCESSION   AR549490
VERSION     AR549490.1  GI:53942665
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1479)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 4621 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .1479
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              341           Length:        1479
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0

Query Match:            100.00%       Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-28 (1-6) x AR549490 (1-1479)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1112 CACAACCACAACCACAAC 1129

RESULT 198
BT004101
LOCUS       BT004101               1479 bp    mRNA    linear   PLN 14-FEB-2003
DEFINITION  Arabidopsis thaliana clone RAFL15-47-E21 (R20969) unknown protein
            (At5g49920) mRNA, complete cds.
ACCESSION   BT004101
VERSION     BT004101.1  GI:28393405
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1479)
  AUTHORS   Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,
            Onodera,C.S., Quach,H.L., Tang,C., Toriumi,M., Wong,C., Wu,H.C.,
            Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,
            Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M.,
            Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P.,
            Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W.,
            Ecker,J.R. and Theologis,A.
  TITLE     Arabidopsis Full Length cDNA Clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1479)
  AUTHORS   Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,
            Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C.,
            Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,
            Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M.,
            Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shinn,P.,
            Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W.,
            Ecker,J.R. and Theologis,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-2003) Plant Gene Expression Center, 800 Buchanan
            Street, Albany, CA 94710, USA
COMMENT     RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis Full-Length cDNA'): Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members carried out the
            sequencing and annotation of the RAFL cDNAs: Yamada,K., Chan,M.M.,
            Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Onodera,C.S.,
            Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C., Yu,G.,
            Yuan,S., Chen,H., Cheuk,R., Jones,T., Kim,C.J., Nguyen,M.,
            Palm,C.J., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Davis,R.W.,
            Ecker,J.R. and Theologis,A.

            Yamada,K. (SSP/PGEC) and Seki,M. (RIKEN GSC) contributed equally to
            this work. Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP/PGEC)
            contributed equally to this work as PIs.

            Annotation based on July 2002 version of the Arabidopsis genome
            submitted to Genbank.
FEATURES             Location/Qualifiers
     source          1. .1479
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /chromosome="5"
                     /clone="RAFL15-47-E21 (R20969)"
                     /ecotype="Columbia"
                     /note="This clone is in a modified pBluescript vector2
                     (lambda PS) as a BamHI/XhoI insert."
```

```
     gene            1. .1479
                     /gene="At5g49920"
     5'UTR           1. .360
                     /gene="At5g49920"
     CDS             361. .1227
                     /gene="At5g49920"
                     /codon_start=1
                     /evidence=experimental
                     /product="unknown protein"
                     /protein_id="AAO42126.1"
                     /db_xref="GI:28393406"
                     /translation="MAEERSPSKVKFMCSFGGRILPRPSDSVLKYVGGETRVVAVSPD
                     ISFSELMKKLTAITENDIVLKYQIIPEDLDALVSVKSDEDVKHMIEEYNRHETPKLRT
                     FLFPANPVVLESQLGPIEPQTIEQRYIEAINSILRTSKSATALRAPIKTRPSFTVSTC
                     SSPKSESPDGYSHEPPETSFQNNYQLSRLYPMHRVHSSPNISHQLQPHSHYHNHNHNA
                     YVQPPTNYLTCRPRLPTPLEIPRGIGWDQSYTTTHNSGGGNGKYGCNDDRRFWGRASS
                     VPQSPRNHGLRL"
     misc_difference 760
                     /gene="At5g49920"
                     /note="compared to genomic sequence resulting in an amino
                     acid sequence difference"
                     /replace="g"
     3'UTR           1228. .1479
                     /gene="At5g49920"
ORIGIN

Alignment Scores:
Pred. No.:                341           Length:         1479
Score:                    42.00         Matches:        6
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              100.00%       Indels:         0
DB:                       15            Gaps:           0

US-09-858-332G-28 (1-6) x BT004101 (1-1479)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        994 CACAACCACAACCACAAT 1011

RESULT 199
AR547211
LOCUS       AR547211                1482 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 2342 from patent US 6747137.
ACCESSION   AR547211
VERSION     AR547211.1  GI:53940386
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1482)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 2342 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .1482
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:                342           Length:         1482
Score:                    42.00         Matches:        6
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              100.00%       Indels:         0
DB:                       6             Gaps:           0

US-09-858-332G-28 (1-6) x AR547211 (1-1482)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        869 CACAACCACAACCACAAC 886

RESULT 200
DDU72746
LOCUS       DDU72746                1486 bp    mRNA    linear   INV 19-MAR-1997
DEFINITION  Dictyostelium discoideum cysteine proteinase (cprG) mRNA, complete
            cds.
ACCESSION   U72746
VERSION     U72746.1  GI:1644501
KEYWORDS    .
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 1486)
  AUTHORS   Ord,T., Adessi,C., Wang,L. and Freeze,H.H.
  TITLE     The cysteine proteinase gene cprG in Dictyostelium discoideum has a
            serine-rich domain that contains GlcNAc-1-P
  JOURNAL   Arch. Biochem. Biophys. 339 (1), 64-72 (1997)
   PUBMED   9056234
REFERENCE   2  (bases 1 to 1486)
  AUTHORS   Ord,T., Adessi,C., Wang,L. and Freeze,H.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-1996) The Burnham Institute, 10901 N. Torrey
            Pines Rd. La Jolla, CA 92093, USA
FEATURES             Location/Qualifiers
     source          1. .1486
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX-2"
                     /db_xref="taxon:44689"
     gene            1. .1486
                     /gene="cprG"
     CDS             9. .1391
                     /gene="cprG"
                     /codon_start=1
                     /product="cysteine proteinase"
                     /protein_id="AAC47482.1"
                     /db_xref="GI:1644502"
                     /translation="MKVLSALCVLLVSVATAKQQLSEVEYRNAFTNWMIAHQRHYSSE
                     EFNGRYNIFKANMDYVNEWNTKGSETVLGLNVFADISNEEYRATYLGTPFDASSLEMT
                     ESDKIFDASAQVDWRTQGAVTPIKNQGQCGGCWSFSTTGATEGAQYLANGKKNLVSLS
                     EQNLIDCSGSYGNNGCEGGLMTLAFEYIINNKGIDTESSYPYTAEDGKKCKFNPKNVA
                     AQLSSYVNVTSGSESDLAAKVTQGPTSVAIDASNQSFQLYVSGIYNEPACSSTQLDHG
                     VLAVGFGTGSGSSGSHGGSQSQSAGSDSQSAGSESSQSESGSQSQSESGSQSQSQSGS
                     QSFSGSLYSGSYSGSQSGSQSGNSGAAVKQTGAGSGSGSGSGSGSGSGSGSVSGSASG
                     SASGSASGSSSGSNSNGGVYPTAGDYWIVKNSWGTSWGMDGYILMTKGNNNQCGIATM
                     ASRPTAVASL"
ORIGIN

Alignment Scores:
Pred. No.:                343           Length:         1486
Score:                    42.00         Matches:        6
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              100.00%       Indels:         0
DB:                       2             Gaps:           0

US-09-858-332G-28 (1-6) x DDU72746 (1-1486)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        985 CACAATCACAATCACAAT 1002

RESULT 201
AR578923
LOCUS       AR578923                1521 bp    DNA     linear   PAT 14-DEC-2004
DEFINITION  Sequence 7 from patent US 6783985.
ACCESSION   AR578923
VERSION     AR578923.1  GI:56581740
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
```

```
            Unclassified.
REFERENCE   1  (bases 1 to 1521)
  AUTHORS   Roemer,T., Jiang,B., Boone,C. and Bussey,H.
  TITLE     Gene disruption methodologies for drug target discovery
  JOURNAL   Patent: US 6783985-A 7 31-AUG-2004;
            Elitra Pharmaceuticals Inc.; San Diego, CA
FEATURES             Location/Qualifiers
     source          1. .1521
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              350           Length:         1521
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     6             Gaps:           0

US-09-858-332G-28 (1-6) x AR578923 (1-1521)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        176 CACAACCACAACCACAAC 193

RESULT 202
AX240769
LOCUS       AX240769               1521 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 7 from Patent WO0160975.
ACCESSION   AX240769
VERSION     AX240769.1  GI:15797705
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1
  AUTHORS   Roemer,T., Jiang,B., Boone,C. and Bussey,H.
  TITLE     Gene disruption methodologies for drug target discovery
  JOURNAL   Patent: WO 0160975-A 7 23-AUG-2001;
            Elitra Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .1521
                     /organism="Candida albicans"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:5476"
ORIGIN

Alignment Scores:
Pred. No.:              350           Length:         1521
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     6             Gaps:           0

US-09-858-332G-28 (1-6) x AX240769 (1-1521)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        176 CACAACCACAACCACAAC 193

RESULT 203
AX488745
LOCUS       AX488745               1521 bp    DNA     linear   PAT 16-AUG-2002
DEFINITION  Sequence 6045 from Patent WO02053728.
ACCESSION   AX488745
VERSION     AX488745.1  GI:22322757
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1
  AUTHORS   Roemer,T., Jiang,B., Boone,C., Bussey,H. and Ohlsen,K.L.
  TITLE     Gene disruption methodologies for drug target discovery
  JOURNAL   Patent: WO 02053728-A 6045 11-JUL-2002;
            Elitra Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .1521
                     /organism="Candida albicans"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:5476"
ORIGIN

Alignment Scores:
Pred. No.:              350           Length:         1521
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     6             Gaps:           0

US-09-858-332G-28 (1-6) x AX488745 (1-1521)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        176 CACAACCACAACCACAAC 193

RESULT 204
AY700082/c
LOCUS       AY700082               1551 bp    DNA     linear   INV 10-FEB-2005
DEFINITION  Plasmodium falciparum DEAD box DNA helicase gene, complete cds.
ACCESSION   AY700082
VERSION     AY700082.1  GI:51556984
KEYWORDS    .
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 1551)
  AUTHORS   Pradhan,A., Chauhan,V.S. and Tuteja,R.
  TITLE     A novel 'DEAD-box' DNA helicase from Plasmodium falciparum is
            homologous to p68
  JOURNAL   Mol. Biochem. Parasitol. 140 (1), 55-60 (2005)
   PUBMED   15694486
REFERENCE   2  (bases 1 to 1551)
  AUTHORS   Tuteja,R. and Pradhan,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUL-2004) Malaria, ICGEB, Aruna Asaf Ali Marg, New
            Delhi, Delhi 110067, India
FEATURES             Location/Qualifiers
     source          1. .1551
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5833"
     mRNA            <1. .>1551
                     /product="DEAD box DNA helicase"
     CDS             1. .1551
                     /note="enzyme"
                     /codon_start=1
                     /product="DEAD box DNA helicase"
                     /protein_id="AAU06262.1"
                     /db_xref="GI:51556985"
                     /translation="MERQNLYNNNYNKCNDQNNNQNSNDNLNNEQTNCLSKEDIQNEL
                     KKNNIYINKDGIIHNIINKFSDVCFHESILNYLNNKFSEPTAIQKITWPIALSGKDLI
                     GVAETGSGKTLAFVLPCFMHILKHKEIIQNNNNNKQTNQNNNLQQEQEGKTNNHHIIN
                     NYNENNDNNENNHNNNNNNNNNNNSYDYDYDINSSERASDTYGLILLPTRELCLQVL
                     DEIKSFEKNLPIKSVAVYGGVPKYYQINNLKKGADIIVATPGRLLDFLENGNINLLKC
                     IYVVIDEADRLLDMGFEKQLRKIMTQVNKNKQLLFLTATWPEQVRKLAYDFCAYDPVK
                     IQIGKNELTANKNIEQNVIISSSIDMKKKLLDWLKENYENNKILIFCDTKRNCDNLGK
                     ELRYHQYNALSIHGDKQQRERDRILNNYKTDRCNILVATDVASRGLDIKNISVVINYD
                     IPNTIEDYIHRIGRTGRAGKKGKSILFFSYDYYMPQKLKFAKELIKLLNKTNQTVPPQ
                     LKEIAYSR"
ORIGIN
```

```
Alignment Scores:
Pred. No.:                357             Length:          1551
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       2               Gaps:            0

US-09-858-332G-28 (1-6) x AY700082 (1-1551)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        574 CATAATCATAATCATAAC 557

RESULT 205
AR548552
LOCUS       AR548552                1575 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 3683 from patent US 6747137.
ACCESSION   AR548552
VERSION     AR548552.1  GI:53941727
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1575)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 3683 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .1575
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:                362             Length:          1575
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       6               Gaps:            0

US-09-858-332G-28 (1-6) x AR548552 (1-1575)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1330 CACAACCACAACCACAAC 1347

RESULT 206
AF005262
LOCUS       AF005262                1610 bp    DNA     linear   PLN 03-SEP-1997
DEFINITION  Cucumis melo randomly amplified polymorphic fragment DNA sequence
            from MR-1, linked to FOM 2.
ACCESSION   AF005262
VERSION     AF005262.1  GI:2352497
KEYWORDS    .
SOURCE      Cucumis melo (muskmelon)
  ORGANISM  Cucumis melo
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Cucurbitales; Cucurbitaceae; Cucumis.
REFERENCE   1  (bases 1 to 1610)
  AUTHORS   Wechter,W.P., Dean,R.A. and Thomas,C.E.
  TITLE     pWEC01 from Cucumis melo
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1610)
  AUTHORS   Wechter,W.P., Dean,R.A. and Thomas,C.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAY-1997) Plant Pathology, Clemson University, 120
            Long Hall, Clemson, SC 29634-0377, USA
FEATURES             Location/Qualifiers
     source          1. .1610
                     /organism="Cucumis melo"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3656"
                     /clone="pWEC01"
ORIGIN

Alignment Scores:
Pred. No.:                370             Length:          1610
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       15              Gaps:            0

US-09-858-332G-28 (1-6) x AF005262 (1-1610)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        258 CACAATCACAATCACAAT 275

RESULT 207
AR359834
LOCUS       AR359834                1614 bp    DNA     linear   PAT 17-AUG-2003
DEFINITION  Sequence 1 from patent US 6593513.
ACCESSION   AR359834
VERSION     AR359834.1  GI:33766633
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1614)
  AUTHORS   Davis,E.L. and Goellner,M.
  TITLE     Endoglucanase gene promoter upregulated by the root-knot nematode
  JOURNAL   Patent: US 6593513-A 1 15-JUL-2003;
            North Carolina State University; Raleigh, NC
FEATURES             Location/Qualifiers
     source          1. .1614
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:                371             Length:          1614
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       6               Gaps:            0

US-09-858-332G-28 (1-6) x AR359834 (1-1614)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        992 CATAATCACAACCACAAC 1009

RESULT 208
CQ608256
LOCUS       CQ608256                1616 bp    DNA     linear   PAT 02-FEB-2004
DEFINITION  Sequence 36014 from Patent WO0171042.
ACCESSION   CQ608256
VERSION     CQ608256.1  GI:41660817
KEYWORDS    .
SOURCE      Drosophila sp.
  ORGANISM  Drosophila sp.
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1
  AUTHORS   Venter,J.C., Adams,M., Li,P.W. and Myers,E.W.
  TITLE     Detection kits, such as nucleic acid arrays, for detecting the
            expression of 10,000 or more Drosophila genes and uses thereof
  JOURNAL   Patent: WO 0171042-A 36014 27-SEP-2001;
```

```
           PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1. .1616
                     /organism="Drosophila sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:7242"
ORIGIN

Alignment Scores:
Pred. No.:              371             Length:         1616
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x CQ608256 (1-1616)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        201 CACAACCATAATCACAAT 218

RESULT 209
AY061640
LOCUS       AY061640               1616 bp    mRNA    linear   PLN 30-JUN-2002
DEFINITION  Prunus persica putative ethylene receptor (ERS1) mRNA, partial cds.
ACCESSION   AY061640
VERSION     AY061640.1  GI:21632786
KEYWORDS    .
SOURCE      Prunus persica (peach)
  ORGANISM  Prunus persica
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Rosales; Rosaceae; Amygdaloideae; Prunus.
REFERENCE   1  (bases 1 to 1616)
  AUTHORS   Bassett,C.L. and Artlip,T.S.
  TITLE     Cloning of a Peach Homolog of the Putative Ethylene Receptor ETR1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1616)
  AUTHORS   Bassett,C.L. and Artlip,T.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-NOV-2001) Appalachian Fruit Research Station,
            USDA-ARS, 45 Wiltshire Road, Kearneysville, WV 25430, USA
FEATURES             Location/Qualifiers
     source          1. .1616
                     /organism="Prunus persica"
                     /mol_type="mRNA"
                     /db_xref="taxon:3760"
     gene            <1. .1616
                     /gene="ERS1"
     CDS             <1. .1481
                     /gene="ERS1"
                     /codon_start=3
                     /product="putative ethylene receptor"
                     /protein_id="AAL30116.1"
                     /db_xref="GI:21632787"
                     /translation="IRSTLDRHTILKTTLVELGRTLGLQECALWMPSRTGMNLQLSHT
                     LNYQIQVGSTVPINLPIINEVFTSSRAIRIPYTCPLARIRPLVGRYVPPEVIALRVPL
                     LNLSNFQINDWPDLSAKSYAIMVLILPTDSTRKWRDHELELVDVVADQVAVALSHAAI
                     LEESMRARDQLMEQNVALDLARREAEMAIHARNDFLAVMNHEMRTPMHAIIALSSLLL
                     ETELISEQRVMIETVLRSSNLLETLINDVLDLSRLEDGSLELDIRKFNLHGVFREVIN
                     LIQPITSVKKLSMSLILSPDLPLCAVGDENRLMQTILNVAGNAVKFTKEGYISIVASV
                     AKPESSRDWRPPEFYPASTDGHFYLRVQVKDSGCGILPQDIPHLFTKFSQPRSGSNRI
                     NDGAGLGLAICRRFINTMGGHIWIESEGIDKGSIVAFIVKLGICDNPCDTTAAHQMAP
                     QALPQPQPQPNYGSGDRIAHRPTVSFRDGGYTSSNPRYQRSI"
ORIGIN

Alignment Scores:
Pred. No.:              371             Length:         1616
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x AY061640 (1-1616)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1363 CACAACCACAACCACAAC 1380

RESULT 210
AR553270/c
LOCUS       AR553270               1632 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 8401 from patent US 6747137.
ACCESSION   AR553270
VERSION     AR553270.1  GI:53946445
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1632)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 8401 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .1632
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              375             Length:         1632
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x AR553270 (1-1632)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         49 CATAATCATAATCATAAT 32

RESULT 211
AR551237
LOCUS       AR551237               1665 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 6368 from patent US 6747137.
ACCESSION   AR551237
VERSION     AR551237.1  GI:53944412
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1665)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 6368 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .1665
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              382             Length:         1665
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0
```

```
US-09-858-332G-28 (1-6) x AR551237 (1-1665)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        221 CACAACCACAACCACAAC 238

RESULT 212
BT008623
LOCUS       BT008623                1691 bp    mRNA    linear   PLN 16-MAY-2003
DEFINITION  Arabidopsis thaliana clone RAFL15-02-N09 (R20142) putative seed
            storage protein (vicilin) (At2g28490) mRNA, complete cds.
ACCESSION   BT008623
VERSION     BT008623.1  GI:30793988
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1691)
  AUTHORS   Yamada,K., Dale,J.M., Hsuan,V.W., Onodera,C.S., Quach,H.,
            Toriumi,M., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H.,
            Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J.,
            Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M.,
            Seki,M., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K.,
            Davis,R.W., Ecker,J.R. and Theologis,A.
  TITLE     Arabidopsis Full Length cDNA Clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1691)
  AUTHORS   Yamada,K., Dale,J.M., Hsuan,V.W., Onodera,C.S., Quach,H.L.,
            Toriumi,M., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H.,
            Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J.,
            Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M.,
            Seki,M., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K.,
            Davis,R.W., Ecker,J.R. and Theologis,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2003) Plant Gene Expression Center, 800 Buchanan
            Street, Albany, CA 94710, USA
COMMENT     RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis Full-Length cDNA'): Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members carried out the
            sequencing and annotation of the RAFL cDNAs: Yamada,K., Dale,J.M.,
            Hsuan,V.W.,   Onodera,C.S., Quach,H.L., Toriumi,M., Wong,C.,
            Wu,H.C., Yu,G., Yuan,S.,  Chen,H., Cheuk,R., Jones,T., Kim,C.J.,
            Nguyen,M., Palm,C.J., Shinn,P.,  Southwick,A., Tripp,M.G., Wu,T.,
            Davis,R.W., Ecker,J.R. and Theologis,A.

            Yamada,K. (SSP/PGEC) and Seki,M. (RIKEN GSC) contributed equally to
            this  work.  Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP/PGEC)
            contributed  equally to this work as PIs.

            Annotation based on July 2002 version of the Arabidopsis genome
            submitted to Genbank.
FEATURES             Location/Qualifiers
     source          1. .1691
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /chromosome="2"
                     /clone="RAFL15-02-N09 (R20142)"
                     /ecotype="Columbia"
                     /note="This clone is in a modified pBluescript vector2
                     (lambda PS) as a BamHI/XhoI insert."
     gene            1. .1691
                     /gene="At2g28490"
     5'UTR           1. .31
                     /gene="At2g28490"
     misc_difference 1. .2
                     /gene="At2g28490"
                     /note="compared to genomic sequence"
                     /replace="ac"
     CDS             32. .1567
                     /gene="At2g28490"
                     /codon_start=1
                     /evidence=experimental
                     /product="putative seed storage protein (vicilin)"
                     /protein_id="AAP40444.1"
                     /db_xref="GI:30793989"
                     /translation="MEKNKRAIGFLLLVVLINGVMMTRSNGYEGEEEWGGAGGGEWGG
                     AEGGGAWGGGGGGGGAWGGEGEGGGEWGGGGEGGGGGRRGWFMMRESRQVIKSEGGEM
                     RVVLSPRGRIIEKPMHIGFLTMEPKTLFVPQYLDSSLLIFIRQGEATLGVICKDEFGE
                     RKLKAGDIYWIPAGSVFYLHNTGLGQRLHVICSIDPTQSLGFETFQPFYIGGGPSSVL
                     AGFDPHTLTSAFNVSLPELQQMMMSQFRGPIVYVTEGPQPQPQSTVWTQFLGLRGEEK
                     HKQLKKLLETKQGSPQDQQYSSGWSWRNIVRSILDLTEEKNKGSGSSECEDSYNIYDK
                     KDKPSFDNKYGWSIALDYDDYKPLKHSGIGVYLVNLTAGAMMAPHMNPTATEYGIVLA
                     GSGEIQVVFPNGTSAMNTRVSVGDVFWIPRYFAFCQIASRTGPFEFVGFTTSAHKNRP
                     QFLVGSNSLLRTLNLTSLSIAFGVDEETMRRFIEAQREAVILPTPAAAPPHVGEMVSD
                     RFV"
     3'UTR           1568. .1691
                     /gene="At2g28490"
     misc_difference 1676
                     /gene="At2g28490"
                     /note="not present in genomic sequence"
ORIGIN

Alignment Scores:
Pred. No.:                388           Length:          1691
Score:                    42.00         Matches:         6
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       15            Gaps:            0

US-09-858-332G-28 (1-6) x BT008623 (1-1691)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        798 CACAACCACAACCACAAT 815

RESULT 213
AK107456
LOCUS       AK107456                1704 bp    mRNA    linear   PLN 24-JUL-2003
DEFINITION  Oryza sativa (japonica cultivar-group) cDNA clone:002-128-C10, full
            insert sequence.
ACCESSION   AK107456
VERSION     AK107456.1  GI:32992665
KEYWORDS    FLI_CDNA; oligo capping.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   The Rice Full-Length cDNA Consortium, National Institute of
            Agrobiological Sciences Rice Full-Length cDNA Project Team:,
            Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K.,
            Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I.,
            Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C.,
            Ohtsuki,K., Shishiki,T., Foundation of Advancement of International
            Science Genome Sequencing & Analysis Group:, Otomo,Y., Murakami,K.,
            Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y.,
            Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M.,
            Narikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Niikura,J.,
            Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Miura,J.,
            Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN:,
            Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuda,S.,
            Hara,A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M.,
            Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Osato,N., Ota,Y.,
            Saito,R., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
            Yoshino,M. and Hayashizaki,Y.
  TITLE     Collection, mapping, and annotation of over 28,000 cDNA clones from
```

```
            japonica rice
  JOURNAL   Science 301 (5631), 376-379 (2003)
   PUBMED   12869764
REFERENCE   2  (bases 1 to 1704)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K.,
            Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiraoka,T.,
            Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K.,
            Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I.,
            Kanagawa,S., Katoh,H., Kawagashira,N., Kawai,J., Kawamata,M.,
            Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M.,
            Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Kurosaki,T., Kusumegi,T., Li,C., Lu,M.,
            Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A.,
            Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakamura,M.,
            Namiki,T., Narikawa,R., Niikura,J., Nishi,K., Nomura,K.,
            Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M., Ooka,H.,
            Osato,N., Ota,Y., Otomo,Y., Ryu,R., Saitoh,H., Sakai,C., Sakai,K.,
            Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Shibata,K.,
            Shinagawa,A., Shiraki,T., Shishiki,T., Sogabe,Y., Sugano,S.,
            Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y.,
            Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A.,
            Toya,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,W.,
            Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and
            Yoshimura,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of
            Agrobiological Sciences, Department of Molecular Genetics, Head of
            Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki
            305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp,
            Tel:81-29-838-7007, Fax:81-29-838-7007)
COMMENT     This clone is one of the 28K full-length cDNA clones from japonica
            rice.
            URL : http://cdna01.dna.affrc.go.jp/cDNA/
            NIAS Rice Full-Length cDNA Project Team: Kikuchi,S., Satoh,K.,
            Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J.,
            Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T.,
            Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T. and
            Yamamoto,M.
            FAIS Genome Sequencing & Analysis Group: Otomo,Y., Iida,Y.,
            Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M., Kobayashi,M.,
            Kodama,T., Kurosaki,T., Kusumegi,T., Lu,M., Masuda,H., Miura,J.,
            Mizuno,K., Narikawa,R., Niikura,J., Oka,M., Ryu,R., Sugano,S.,
            Sugiyama,A., Suzuki,Y., Tsunoda,Y., Ueda,M., Xie,Q., Yokomizo,S.,
            Yoshimura,A., Matsubara,K. and Murakami,K.
            Genome Exploration Research Group in Riken Genomic Sciences Center
            and Genome Science Laboratory in Riken: Adachi,J., Aizawa,K.,
            Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hanagaki,T.,
            Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K.,
            Hiraoka,T., Hori,F., Iida,J., Imamura,K., Imotani,K., Ishii,Y.,
            Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J.,
            Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Osato,N.,
            Ota,Y., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H.,
            Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F.,
            Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K.,
            Yasunishi,A. and Hayashizaki,Y.
FEATURES              Location/Qualifiers
     source           1. .1704
                      /organism="Oryza sativa (japonica cultivar-group)"
                      /mol_type="mRNA"
                      /cultivar="Nipponbare"
                      /db_xref="taxon:39947"
                      /clone="002-128-C10"
ORIGIN

Alignment Scores:
Pred. No.:                391          Length:        1704
Score:                    42.00        Matches:       6
Percent Similarity:       100.00%      Conservative:  0
Best Local Similarity:    100.00%      Mismatches:    0

Query Match:              100.00%      Indels:        0
DB:                       15           Gaps:          0

US-09-858-332G-28 (1-6) x AK107456 (1-1704)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        307 CACAACCACAACCACAAC 324

RESULT 214
SWP789G/c
LOCUS       SWP789G                 1716 bp    DNA     linear   VRL 03-AUG-1993
DEFINITION  Swinepox virus SwF7, SwF8, SwF8a and SwF9 genes.
ACCESSION   M59931
VERSION     M59931.1  GI:335128
KEYWORDS    thymidine kinase.
SOURCE      Swinepox virus
  ORGANISM  Swinepox virus
            Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
            Suipoxvirus.
REFERENCE   1  (bases 1 to 1716)
  AUTHORS   Schnitzlein,W.M. and Tripathy,D.N.
  TITLE     Identification and nucleotide sequence of the thymidine kinase gene
            of swinepox virus
  JOURNAL   Virology 181 (2), 727-732 (1991)
   PUBMED   1840707
COMMENT     Original source text: Swinepox virus DNA.
FEATURES              Location/Qualifiers
     source           1. .1716
                      /organism="Swinepox virus"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:10276"
     gene             1. .288
                      /gene="SwF7"
     CDS              <1. .288
                      /gene="SwF7"
                      /codon_start=1
                      /protein_id="AAA47890.1"
                      /db_xref="GI:335129"
                      /translation="KLESIGYLYEPLSEEYRRVIDFSDMKNLRSMFNKITIHVSDKCI
                      QVNKGYLSDFVTSLIRLSDSDINTYDSFDITYIDPRRHITWNNILSILNEK"
     gene             312. .845
                      /gene="SwF8"
     CDS              312. .845
                      /gene="SwF8"
                      /codon_start=1
                      /product="thymidine kinase"
                      /protein_id="AAA47891.1"
                      /db_xref="GI:335130"
                      /translation="MSCGFIHLILGPMFSGKSTELIRLVNRYQIATYNCRVIKYSKDN
                      RYGNDAVYTHDKCYISAVSTDSLFDIKDTLDDVDIVGIDEGQFFNDIVEFCEYIANKG
                      KIVIVAALDGTYERKPFGNILNLIPLSEKVTKLNAICMICHRDASFSKRLSDEKEIEL
                      IGGKEKYLSVCRSCYLT"
     gene             900. .1457
                      /gene="SwF8a"
     CDS              900. .1457
                      /gene="SwF8a"
                      /codon_start=1
                      /protein_id="AAA47892.1"
                      /db_xref="GI:335131"
                      /translation="MGITHELDIFLVSEDIAMKHVELHKGNSYGCVLNIKSSCRKQMK
                      IIFVLKPDWTEIDSIKPIQMEADGIYIDVTIIKQSFYEVIYCSNFTIYGDSVINIFSD
                      IDRCTKEKYPIININIDKKKYEIIESGYTYIYIQSPIDKEDKLTYMRDDYVDEYNSSD
                      DDCIDYYDCDYDDDDNNDAEDEDEE"
     gene             1510. .1716
                      /gene="SwF9"
     CDS              1510. .>1716
                      /gene="SwF9"
                      /codon_start=1
                      /protein_id="AAA47893.1"
                      /db_xref="GI:335132"
                      /translation="MEPLSIDKPFMYFDEINNELEYDPDSANEKHKKFPYQGQLKLLL
                      CELFFLSKLQRHGILDGSTIVYIGS"
```

```
ORIGIN

Alignment Scores:
Pred. No.:                394             Length:         1716
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       13              Gaps:           0

US-09-858-332G-28 (1-6) x SWP789G (1-1716)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1413 CATAATCACAATCATAAT 1396

RESULT 215
AR449702/c
LOCUS       AR449702               1731 bp    DNA     linear   PAT 20-FEB-2004
DEFINITION  Sequence 366 from patent US 6673910.
ACCESSION   AR449702
VERSION     AR449702.1  GI:42678759
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1731)
  AUTHORS   Breton,G.L.
  TITLE     Nucleic acid and amino acid sequences relating to M. catarrhalis
            for diagnostics and therapeutics
  JOURNAL   Patent: US 6673910-A 366 06-JAN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .1731
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:                397             Length:         1731
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x AR449702 (1-1731)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1344 CACAATCATAATCATAAT 1327

RESULT 216
AR096201
LOCUS       AR096201               1770 bp    DNA     linear   PAT 08-SEP-2000
DEFINITION  Sequence 1 from patent US 6005092.
ACCESSION   AR096201
VERSION     AR096201.1  GI:10024789
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1770)
  AUTHORS   Shoseyov,O. and Shani,Z.
  TITLE     Arabidopsis thaliana endo-1,4-.beta.-glucanase gene and promoter
  JOURNAL   Patent: US 6005092-A 1 21-DEC-1999;
FEATURES             Location/Qualifiers
     source          1. .1770
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

Alignment Scores:
Pred. No.:                405             Length:         1770
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x AR096201 (1-1770)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        996 CATAATCACAACCACAAC 1013

RESULT 217
BD080270
LOCUS       BD080270               1770 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Transgenic plant with altered morphology and
            endo-1,4-beta-glucanase gene, promoter and protein isolated from
            Arabidopsis thaliana.
ACCESSION   BD080270
VERSION     BD080270.1  GI:22625873
KEYWORDS    JP 2001512672-A/1.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1770)
  AUTHORS   Shoseyov,O., Shani,Z. and Shpiegl,E.
  TITLE     Transgenic plant with altered morphology and
            endo-1,4-beta-glucanase gene, promoter and protein isolated from
            Arabidopsis thaliana
  JOURNAL   Patent: JP 2001512672-A 1 28-AUG-2001;
            YISSUM RESEARCH DEVELOPMENT CO OF THE HEBREW UNIVERSITY OF
            JERUSALEM SECRETARY OF THE DEPARTMENT OF HEAL
COMMENT     OS   Nucleic acid
            PN   JP 2001512672-A/1
            PD   28-AUG-2001
            PF   26-JUL-1998 JP 2000506315
            PR   27-JUL-1997 IL       121404,13-JAN-1998 US    09/006632 PR
            13-JAN-1998 US    09/006636
            PI   ODED SHOSEYOV,ZIV SHANI,ETAI SHPIEGL
            PC   A01H5/00,C12N5/10,C12N9/42,C12N15/09,C12N5/00,C12N15/00 CC
            Transgenic plant with altered morphology and endo-1,4-beta- CC
                 glucanase gene,
            CC   promoter and protein isolated from Arabidopsis thaliana FH
            Key             Location/Qualifiers
            FT   source          1. .1770
            FT                   /organism='Nucleic acid'.
FEATURES             Location/Qualifiers
     source          1. .1770
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN

Alignment Scores:
Pred. No.:                405             Length:         1770
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x BD080270 (1-1770)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        996 CATAATCACAACCACAAC 1013

RESULT 218
AR262064
LOCUS       AR262064               1770 bp    DNA     linear   PAT 29-JAN-2003
DEFINITION  Sequence 1 from patent US 6323023.
ACCESSION   AR262064
```

```
VERSION     AR262064.1  GI:28073411
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1770)
  AUTHORS   Shoseyov,O. and Shani,Z.
  TITLE     Vectors containing nucleic acids coding for Arabidopsis thaliana
            endo-1,4-.beta.-glucanase secretion signal peptide
  JOURNAL   Patent: US 6323023-A 1 27-NOV-2001;
            Yissum Research Development Co., Ltd.; Jerusalem;
            ILX;
FEATURES             Location/Qualifiers
     source          1. .1770
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              405          Length:        1770
Score:                  42.00        Matches:       6
Percent Similarity:     100.00%      Conservative:  0
Best Local Similarity:  100.00%      Mismatches:    0
Query Match:            100.00%      Indels:        0
DB:                     6            Gaps:          0

US-09-858-332G-28 (1-6) x AR262064 (1-1770)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        996 CATAATCACAACCACAAC 1013

RESULT 219
DMRSIG1
LOCUS       DMRSIG1                 1798 bp    DNA     linear   INV 11-MAY-1992
DEFINITION  D.melanogaster regulatory signal sequence (clone scs).
ACCESSION   X63731
VERSION     X63731.1  GI:8498
KEYWORDS    boundary function; regulatory signal.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1
  AUTHORS   Farkas,G. and Udvardy,A.
  TITLE     Sequence of scs and scs' Drosophila DNA fragments with boundary
            function in the control of gene expression
  JOURNAL   Nucleic Acids Res. 20 (10), 2604 (1992)
   PUBMED   1598225
REFERENCE   2  (bases 1 to 1798)
  AUTHORS   Udvardy,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-FEB-1992) A. Udvardy, Institute of Biochemistry,
            Biological Research Center, 6701 Szeged, P.O.Box 521, HUNGARY
COMMENT     See also X63732.
FEATURES             Location/Qualifiers
     source          1. .1798
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /strain="Canton S"
                     /db_xref="taxon:7227"
                     /chromosome="3, 87A7 band"
                     /clone="lambda scs"
                     /clone_lib="lambda EMBL3"
                     /dev_stage="embryo"
     misc_feature    1. .1798
                     /note="regulatory siganl"
                     /evidence=experimental
ORIGIN

Alignment Scores:
Pred. No.:              412          Length:        1798
Score:                  42.00        Matches:       6
Percent Similarity:     100.00%      Conservative:  0
Best Local Similarity:  100.00%      Mismatches:    0
Query Match:            100.00%      Indels:        0
DB:                     2            Gaps:          0

US-09-858-332G-28 (1-6) x DMRSIG1 (1-1798)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1626 CACAACCATAATCACAAT 1643

RESULT 220
AX823825
LOCUS       AX823825                1798 bp    DNA     linear   PAT 11-DEC-2003
DEFINITION  Sequence 3 from Patent WO03070958.
ACCESSION   AX823825
VERSION     AX823825.1  GI:39750144
KEYWORDS    .
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1
  AUTHORS   Yao,S. and Ellis,J.
  TITLE     Retroviral gene therapy vectors including insulator elements to
            provide high levels of gene expression
  JOURNAL   Patent: WO 03070958-A 3 28-AUG-2003;
            THE HOSPITAL FOR SICK CHILDREN (CA)
FEATURES             Location/Qualifiers
     source          1. .1798
                     /organism="Drosophila melanogaster"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:7227"
ORIGIN

Alignment Scores:
Pred. No.:              412          Length:        1798
Score:                  42.00        Matches:       6
Percent Similarity:     100.00%      Conservative:  0
Best Local Similarity:  100.00%      Mismatches:    0
Query Match:            100.00%      Indels:        0
DB:                     6            Gaps:          0

US-09-858-332G-28 (1-6) x AX823825 (1-1798)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1626 CACAACCATAATCACAAT 1643

RESULT 221
AR545053
LOCUS       AR545053                1806 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 184 from patent US 6747137.
ACCESSION   AR545053
VERSION     AR545053.1  GI:53938228
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1806)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 184 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .1806
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN
```

```
Alignment Scores:
Pred. No.:                413             Length:         1806
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x AR545053 (1-1806)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1150 CATAACCACAACCATAAC 1167

RESULT 222
AR447922/c
LOCUS       AR447922                1813 bp    DNA     linear   PAT 20-FEB-2004
DEFINITION  Sequence 558 from patent US 6673549.
ACCESSION   AR447922
VERSION     AR447922.1  GI:42676246
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1813)
  AUTHORS   Furness,L.M. and Buchbinder,J.L.
  TITLE     Genes expressed in C3A liver cell cultures treated with steroids
  JOURNAL   Patent: US 6673549-A 558 06-JAN-2004;
            Incyte Corporation; Palo Alto, CA
FEATURES             Location/Qualifiers
     source          1. .1813
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:                415             Length:         1813
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x AR447922 (1-1813)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1608 CATAATCATAATCATAAT 1591

RESULT 223
AY128494
LOCUS       AY128494                1853 bp    mRNA    linear   INV 08-JUL-2002
DEFINITION  Drosophila melanogaster RH42690 full insert cDNA.
ACCESSION   AY128494
VERSION     AY128494.1  GI:21711792
KEYWORDS    FLI_CDNA.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1853)
  AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J.,
            Champe,M., Chavez,C., Dorsett,V., Dresnek,D., Farfan,D., Frise,E.,
            George,R., Gonzalez,M., Guarin,H., Kronmiller,B., Li,P., Liao,G.,
            Miranda,A., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
            Patel,S., Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Rubin,G.M.
            and Celniker,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-2002) Berkeley Drosophila Genome Project,
            Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, USA
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory
            Berkeley, CA 94720
            This clone was sequenced as part of a high-throughput process to
            sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
            Science 2000). The sequence has been subjected to integrity checks
            for sequence accuracy, presence of a polyA tail and contiguity
            within 100 kb in the genome. Thus we believe the sequence to
            reflect accurately this particular cDNA clone. However, there are
            artifacts associated with the generation of cDNA clones that may
            have not been detected in our initial analyses such as internal
            priming, priming from contaminating genomic DNA, retained introns
            due to reverse transcription of unspliced precursor RNAs, and
            reverse transcriptase errors that result in single base changes.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our Web site
            (http://fruitfly.berkeley.edu) or send email to
            cdna@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1. .1853
                     /organism="Drosophila melanogaster"
                     /mol_type="mRNA"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
     gene            1. .1853
                     /gene="CG3198"
                     /db_xref="FLYBASE:FBgn0029887"
     CDS             233. .1549
                     /gene="CG3198"
                     /note="Longest ORF"
                     /codon_start=1
                     /product="RH42690p"
                     /protein_id="AAM75087.1"
                     /db_xref="GI:21711793"
                     /db_xref="FLYBASE:FBgn0029887"
                     /translation="MVDAARQMLDELMGRNRNLHPSEAGAKVNWEDPEFCQFYNVKFC
                     PHDLFINTRADLGPCARIHDEEARHLYEDARPSQRKRQYEDEFLRFCNVMLHDVDRKI
                     QKGKQRLLLMQRDQPNVPAPLSRHQEQLANLTARINKLLSEAEEAGIRGDVDQAQDLM
                     TVCEELKEEKEQLVQQYEAHHKAISSISLTKSAEDELQHHPRSSPVTNPSEESPTTPL
                     PAATTLDEAPDAAPSSAQAAVTTTDDAGEAAGAASEDKTSEKTDSKAAGWSHDAMPEK
                     QMKVCEICGAFLIVGDAQQRIEDHLMGKQHLGYSKLRNAVAEINEARQKEREQEERRR
                     REGRVQRFHSYDSRREPRSEYRERSRDYVQNRQPSDRRHHHHKSHHSSHHSYSRSGGG
                     GGGVGVGGGGGGSSRSNRSHHNHNHNHYRHDSRERHCRSRSRSRY"
ORIGIN

Alignment Scores:
Pred. No.:                424             Length:         1853
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       2               Gaps:           0

US-09-858-332G-28 (1-6) x AY128494 (1-1853)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1472 CATAATCACAACCATAAT 1489

RESULT 224
I34914
LOCUS       I34914                  1897 bp    DNA     linear   PAT 13-MAY-1997
DEFINITION  Sequence 1 from patent US 5599701.
ACCESSION   I34914
VERSION     I34914.1  GI:2087882
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1897)
  AUTHORS   Henry,S.A. and White,M.J.
  TITLE     Inositol-excreting yeast
  JOURNAL   Patent: US 5599701-A 1 04-FEB-1997;
```

```
FEATURES             Location/Qualifiers
     source          1. .1897
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

Alignment Scores:
Pred. No.:              433           Length:         1897
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     6             Gaps:           0

US-09-858-332G-28 (1-6) x I34914 (1-1897)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1730 CATAATCATAATCATAAT 1747

RESULT 225
YSCOPI1
LOCUS       YSCOPI1                 1897 bp    DNA     linear   PLN 27-APR-1993
DEFINITION  Saccharomyces cerevisiae negative phospholipid biosynthesis
            regulator (OPI1) gene, complete cds.
ACCESSION   M57383 J05727
VERSION     M57383.1  GI:172067
KEYWORDS    negative phospholipid biosynthesis regulator.
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 1897)
  AUTHORS   White,M.J., Hirsch,J.P. and Henry,S.A.
  TITLE     The OPI1 gene of Saccharomyces cerevisiae, a negative regulator of
            phospholipid biosynthesis, encodes a protein containing
            polyglutamine tracts and a leucine zipper
  JOURNAL   J. Biol. Chem. 266 (2), 863-872 (1991)
   PUBMED   1985968
COMMENT     Original source text: S.cerevisiae DNA.
FEATURES             Location/Qualifiers
     source          1. .1897
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4932"
     gene            439. .1653
                     /gene="OPI1"
     CDS             439. .1653
                     /gene="OPI1"
                     /codon_start=1
                     /product="negative phospholipid biosynthesis regulator"
                     /protein_id="AAA34828.1"
                     /db_xref="GI:172068"
                     /translation="MSENQRLGLSEEEVEAAEVLGVLKQSCRQKSQPSEDVSQADKMP
                     ASESSTTPLNILDRVSNKIISNVVTFYDEINTNKRPLKSIGRLLDDDDDEHDDYDYND
                     DEFFTNKRQKLSRAIAKGKDNLKEYKLNMSIESKKRLVTCLHLLKLANKQLSDKISCL
                     QDLVEKEQVHPLHKQDGNARTTTGAGEDETSSDEDDDDEEFFDASEQVNASEQSIVVK
                     MEVVGTVKKVYSLISKFTANSLPEPARSQVRESLLNLPTNWFDSVHSTSLPHHASFHY
                     ANCEEQKVEQQQQQQQQQQQQQLLQQQLLQQQQQKRNKDGDDSASPSSSVTANGKVLI
                     LAKESLEMVRNVMGVVDSTLGKAEEWVKQKQEVKEMIRERFLQQQQQYRQQQQKDGNY
                     VKPSQDNVDSKD"
ORIGIN

Alignment Scores:
Pred. No.:              433           Length:         1897
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     15            Gaps:           0

US-09-858-332G-28 (1-6) x YSCOPI1 (1-1897)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1730 CATAATCATAATCATAAT 1747

RESULT 226
AX536592
LOCUS       AX536592                1898 bp    DNA     linear   PAT 22-NOV-2002
DEFINITION  Sequence 193 from Patent WO02064766.
ACCESSION   AX536592
VERSION     AX536592.1  GI:25262997
KEYWORDS    .
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1
  AUTHORS   Contreras,R.H., Eberhardt,I., Luyten,W.H. and Reekmans,R.J.
  TITLE     Bax-responsive genes for drug target identification in yeast and
            fungi
  JOURNAL   Patent: WO 02064766-A 193 22-AUG-2002;
            JANSSEN PHARMACEUTICA N.V. (BE)
FEATURES             Location/Qualifiers
     source          1. .1898
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:4932"
ORIGIN

Alignment Scores:
Pred. No.:              434           Length:         1898
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     6             Gaps:           0

US-09-858-332G-28 (1-6) x AX536592 (1-1898)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        195 CATAATCATAATCATAAT 212

RESULT 227
AY059751
LOCUS       AY059751                1900 bp    mRNA    linear   PLN 18-SEP-2002
DEFINITION  Arabidopsis thaliana unknown protein (At1g78420) mRNA, complete
            cds.
ACCESSION   AY059751
VERSION     AY059751.1  GI:16604612
KEYWORDS    FLI_CDNA.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1900)
  AUTHORS   Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,
            Goldsmith,A.D., Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L.,
            Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J.,
            Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B.,
            Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T.,
            Satou,M., Seki,M., Shinn,P., Southwick,A., Shinozaki,K.,
            Davis,R.W., Ecker,J.R. and Theologis,A.
  TITLE     Arabidopsis Full Length cDNA Clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1900)
  AUTHORS   Yamada,K., Banh,J., Banno,F., Chang,E., Dale,J.M., Goldsmith,A.D.,
            Lee,J.M., Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wu,H.C.,
            Yamamura,Y., Yu,G., Yu,S., Bowser,L., Carninci,P., Chen,H.,
            Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A.,
            Karlin-Neumann,G., Kawai,J., Kim,C., Koesema,E., Lam,B., Lin,J.,
            Meyers,M.C., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J.,
            Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tracy,S.E.,
```

```
            Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-OCT-2001) Plant Gene Expression Center, 800 Buchanan
            Street, Albany, CA 94710, USA
COMMENT     RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis Full-Length cDNA'): Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members carried out the
            sequencing and annotation of the RAFL cDNAs: Yamada,K., Banh,J.,
            Banno,F., Chang,E., Dale,J.M., Goldsmith,A.D., Lee,J.M.,
            Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wu,H.C.,
            Yamamura,Y., Yu,G., Yu,S., Bowser,L., Chen,H., Cheuk,R., Jones,T.,
            Karlin-Neumann,G., Kim,C., Koesema,E., Lam,B., Lin,J., Meyers,M.C.,
            Miranda,M., Nguyen,M., Palm,C.J., Shinn,P., Southwick,A.,
            Tracy,S.E., Davis,R.W., Ecker,J.R. and Theologis,A.

            Yamada,K. (SSP/PGEC) and Seki,M. (RIKEN GSC) contributed equally to
            this work.  Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP/PGEC)
            contributed equally to this work as PIs.

            Annotation is based on the January 2002 version of the Arabidopsis
            genome submitted to GenBank.
FEATURES             Location/Qualifiers
     source          1. .1900
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /chromosome="1"
                     /clone="RAFL07-07-N11 (R10670)"
                     /ecotype="Columbia"
                     /note="This clone is in a modified pBluescript vector
                     (FLC-1) as a BamHI/XhoI insert."
     gene            1. .1900
                     /gene="At1g78420"
     5'UTR           1. .463
                     /gene="At1g78420"
     misc_difference 89
                     /gene="At1g78420"
                     /note="not present in genomic sequence"
     CDS             464. .1669
                     /gene="At1g78420"
                     /codon_start=1
                     /evidence=experimental
                     /product="unknown protein"
                     /protein_id="AAL24099.1"
                     /db_xref="GI:16604613"
                     /translation="MGNKLGRKRQVVEERYTKPQGLYVNKDVDVKKLRKLIVESKLAP
                     CYPGDDESCHDLEECPICFLYYPSLNRSRCCMKSICTECFLQMKNPNSARPTQCPFCK
                     TPNYAVEYRGVKSKEEKGIEQVEEQRVIEAKIRMRQKEMQDDEEKMQKRLESCSSSTS
                     AMTGEMEYGSTSAISYNSLMDDGEIAPSQNASVVRQHSRPRGNREDEVDVDLEELMVM
                     EAIWLSVQETGTQRNSASGEITSSRQYVTDNHSYVSSPPRVTPIVEPATPSSSSGGLS
                     CAISALAERQMVGESSSHNHNHNVNVSSYSMLPGNCDSYYDIEQEVDGIDNHHHHRHH
                     YEMGETGSSNSYVSSYMTGEGFHNFPPPPPLVIVPESFEEQMMMAMAVSMAEVHATTT
                     CAPTEVTWQ"
     3'UTR           1670. .1900
                     /gene="At1g78420"
     misc_difference 1884
                     /gene="At1g78420"
                     /note="not present in genomic sequence"
ORIGIN

Alignment Scores:
Pred. No.:                434             Length:         1900
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       15              Gaps:           0

US-09-858-332G-28 (1-6) x AY059751 (1-1900)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1343 CACAATCATAATCACAAT 1360

RESULT 228
CQ582981
LOCUS       CQ582981               1929 bp    DNA     linear   PAT 02-FEB-2004
DEFINITION  Sequence 10739 from Patent WO0171042.
ACCESSION   CQ582981
VERSION     CQ582981.1  GI:41643967
KEYWORDS    .
SOURCE      Drosophila sp.
  ORGANISM  Drosophila sp.
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1
  AUTHORS   Venter,J.C., Adams,M., Li,P.W. and Myers,E.W.
  TITLE     Detection kits, such as nucleic acid arrays, for detecting the
            expression of 10,000 or more Drosophila genes and uses thereof
  JOURNAL   Patent: WO 0171042-A 10739 27-SEP-2001;
            PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1. .1929
                     /organism="Drosophila sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:7242"
ORIGIN

Alignment Scores:
Pred. No.:                440             Length:         1929
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x CQ582981 (1-1929)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1213 CACAACCACAACCACAAC 1230

RESULT 229
BC029741
LOCUS       BC029741               1949 bp    mRNA    linear   ROD 07-OCT-2003
DEFINITION  Mus musculus deleted in polyposis 1-like 1, mRNA (cDNA clone
            MGC:25862 IMAGE:4195889), complete cds.
ACCESSION   BC029741
VERSION     BC029741.1  GI:20988223
KEYWORDS    MGC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1949)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
```

```
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
REFERENCE   2  (bases 1 to 1949)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
            Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
            A.N., Gibbs, R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 30 Row: p Column: 24
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 21314829.
FEATURES             Location/Qualifiers
     source          1. .1949
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="MGC:25862 IMAGE:4195889"
                     /tissue_type="Liver, normal. 5 month old male mouse."
                     /clone_lib="NCI_CGAP_Li9"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     gene            1. .1949
                     /gene="Dpl11"
                     /db_xref="GeneID:70335"
                     /db_xref="MGI:1917585"
     CDS             74. .598
                     /gene="Dpl11"
                     /codon_start=1
                     /product="Dpl11 protein"
                     /protein_id="AAH29741.1"
                     /db_xref="GI:20988224"
                     /db_xref="GeneID:70335"
                     /db_xref="MGI:1917585"
                     /translation="MDGLRQRFERFLEQKNVATEALGALEARTGVEKRYLAAGALALL
                     GLYLLFGYGASLLCNVIGFVYPAYASVKAIESPSKEDDTVWLTYWVVYALFGLVEFFS
                     DLLLFWFPFYYAGKCAFLLFCMTPGPWNGALLLYHRVIRPLFLKHHMALDSAASQLSG
                     RALDLAAGITRDAK"
     misc_feature    245. .505
                     /gene="Dpl11"
                     /note="TB2_DP1_HVA22; Region: TB2/DP1, HVA22 family. This
                     family includes members from a wide variety of eukaryotes.
                     It includes the TB2/DP1 (deleted in polyposis) protein,
                     which in humans is deleted in severe forms of familial
                     adenomatous polyposis, an autosomal dominant oncological
                     inherited disease. The family also includes the plant
                     protein of known similarity to TB2/DP1, the HVA22 abscisic
                     acid-induced protein, which is thought to be a regulatory
                     protein"
                     /db_xref="CDD:pfam03134"

ORIGIN

Alignment Scores:
Pred. No.:                445             Length:          1949
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       9               Gaps:            0

US-09-858-332G-28 (1-6) x BC029741 (1-1949)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        776 CACAACCACAACCACAAC 793

RESULT 230
AR549509
LOCUS       AR549509               1953 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 4640 from patent US 6747137.
ACCESSION   AR549509
VERSION     AR549509.1  GI:53942684
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1953)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 4640 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .1953
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:                446             Length:          1953
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       6               Gaps:            0

US-09-858-332G-28 (1-6) x AR549509 (1-1953)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1691 CACAATCACAATCACAAT 1708

RESULT 231
CQ610475
LOCUS       CQ610475               1958 bp    DNA     linear   PAT 02-FEB-2004
DEFINITION  Sequence 38233 from Patent WO0171042.
ACCESSION   CQ610475
VERSION     CQ610475.1  GI:41662296
KEYWORDS    .
SOURCE      Drosophila sp.
  ORGANISM  Drosophila sp.
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1
  AUTHORS   Venter,J.C., Adams,M., Li,P.W. and Myers,E.W.
  TITLE     Detection kits, such as nucleic acid arrays, for detecting the
            expression of 10,000 or more Drosophila genes and uses thereof
  JOURNAL   Patent: WO 0171042-A 38233 27-SEP-2001;
            PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1. .1958
                     /organism="Drosophila sp."
```

```
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:7242"
ORIGIN

Alignment Scores:
Pred. No.:                447             Length:         1958
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x CQ610475 (1-1958)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        660 CACAACCACAACCACAAT 677

RESULT 232
AR557473
LOCUS       AR557473                1995 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 12604 from patent US 6747137.
ACCESSION   AR557473
VERSION     AR557473.1  GI:53950648
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1995)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 12604 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .1995
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:                455             Length:         1995
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x AR557473 (1-1995)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1304 CACAACCACAACCACAAC 1321

RESULT 233
AX509064
LOCUS       AX509064                2000 bp    DNA     linear   PAT 27-SEP-2002
DEFINITION  Sequence 3759 from Patent WO0216655.
ACCESSION   AX509064
VERSION     AX509064.1  GI:23390301
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Harper,J.F., Kreps,J., Wang,X. and Zhu,T.
  TITLE     Stress-regulated genes of plants, transgenic plants containing
            same, and methods of use
  JOURNAL   Patent: WO 0216655-A 3759 28-FEB-2002;
            The Scripps Research Institute (US) ; Syngenta Participations AG
            (CH)
FEATURES             Location/Qualifiers
     source          1. .2000
                     /organism="Arabidopsis thaliana"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:3702"
ORIGIN

Alignment Scores:
Pred. No.:                456             Length:         2000
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x AX509064 (1-2000)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1332 CATAATCACAACCACAAC 1349

RESULT 234
AX509136/c
LOCUS       AX509136                2000 bp    DNA     linear   PAT 27-SEP-2002
DEFINITION  Sequence 3831 from Patent WO0216655.
ACCESSION   AX509136
VERSION     AX509136.1  GI:23390373
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Harper,J.F., Kreps,J., Wang,X. and Zhu,T.
  TITLE     Stress-regulated genes of plants, transgenic plants containing
            same, and methods of use
  JOURNAL   Patent: WO 0216655-A 3831 28-FEB-2002;
            The Scripps Research Institute (US) ; Syngenta Participations AG
            (CH)
FEATURES             Location/Qualifiers
     source          1. .2000
                     /organism="Arabidopsis thaliana"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:3702"
ORIGIN

Alignment Scores:
Pred. No.:                456             Length:         2000
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x AX509136 (1-2000)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        525 CATAATCATAACCATAAC 508

RESULT 235
AB039933
LOCUS       AB039933                2011 bp    mRNA    linear   ROD 25-FEB-2005
DEFINITION  Mus musculus Dp111 mRNA for polyposis locus protein 1-like 1 (TB2
            protein-like 1), complete cds.
ACCESSION   AB039933
VERSION     AB039933.1  GI:7573634
KEYWORDS    polyposis locus protein 1-like 1.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
```

```
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Sato,H., Tomita,H., Nakazawa,T., Wakana,S. and Tamai,M.
  TITLE     Deleted in Polyposis 1-like 1 Gene (Dpl11): A Novel Gene Richly
            Expressed in Retinal Ganglion Cells
  JOURNAL   Invest. Ophthalmol. Vis. Sci. 46, 791-796 (2005)
REFERENCE   2  (bases 1 to 2011)
  AUTHORS   Sato,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-2000) Hajime Sato, Tohoku University School of
            Medicine, Department of Ophthalmology; Seiryo-machi 1-1, Sendai,
            Miyagi, 980-8574, Japan (E-mail:hasato@oph.med.tohoku.ac.jp,
            Tel:81-22-717-7294, Fax:81-22-717-7298)
FEATURES             Location/Qualifiers
     source          1. .2011
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL6"
                     /db_xref="taxon:10090"
                     /chromosome="10"
                     /clone="pSK138"
                     /tissue_type="retina"
                     /dev_stage="adult"
                     /note="vector:pBluescriptSK(-)"
     gene            1. .2011
                     /gene="Dpl11"
     CDS             62. .667
                     /gene="Dpl11"
                     /standard_name="deleted in polyposis 1-like 1"
                     /codon_start=1
                     /product="polyposis locus protein 1-like 1 (TB2
                     protein-like 1)"
                     /protein_id="BAA94544.1"
                     /db_xref="GI:7573635"
                     /translation="MDGLRQRFERFLEQKNVATEALGALEARTGVEKRYLAAGALALL
                     GLYLLFGYGASLLCNVIGFVYPAYASVKAIESPSKEDDTVWLTYWVVYALFGLVEFFS
                     DLLLFWFPFYYAGKCAFLLFCMTPGPWNGALLLYHRVIRPLFLKHHMALDSAASQLSG
                     RALDLAAGITRDVLQALARGRALVTPASTSEPPAALELDPK"
     polyA_signal    1968. .1973
                     /gene="Dpl11"
ORIGIN

Alignment Scores:
Pred. No.:              458             Length:         2011
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     9               Gaps:           0

US-09-858-332G-28 (1-6) x AB039933 (1-2011)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        845 CACAACCACAACCACAAC 862

RESULT 236
DROORFB/c
LOCUS       DROORFB                 2012 bp    DNA     linear   INV 24-FEB-1994
DEFINITION  D.virilis homeobox protein (antennapedia) gene, 3' end.
ACCESSION   M95826
VERSION     M95826.1  GI:158022
KEYWORDS    antennapedia protein; homeobox protein.
SOURCE      Drosophila virilis
  ORGANISM  Drosophila virilis
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 2012)
  AUTHORS   Hooper,J.E., Perez-Alonso,M., Bermingham,J.R., Prout,M.,
            Rocklein,B.A., Wagenbach,M., Edstrom,J.-E., de Frutos,R. and
            Scott,M.P.
  TITLE     Comparative studies of Drosophila antennapedia genes
  JOURNAL   Unpublished (1992)
COMMENT     Original source text: Drosophila virilis DNA.
FEATURES             Location/Qualifiers
     source          1. .2012
                     /organism="Drosophila virilis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7244"
                     /germline
     gene            1. .2012
                     /gene="antennapedia"
     intron          <1. .243
                     /gene="antennapedia"
     exon            244. .2012
                     /gene="antennapedia"
     CDS             <244. .494
                     /gene="antennapedia"
                     /function="DNA binding"
                     /note="putative"
                     /codon_start=3
                     /product="homeobox protein"
                     /protein_id="AAA28737.1"
                     /db_xref="GI:158023"
                     /translation="RKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQ
                     IKIWFQNRRMKWKKENKTKGEPGSGGEGDEITPPNSPQ"
     3'UTR           492. .2012
                     /gene="antennapedia"
ORIGIN

Alignment Scores:
Pred. No.:              459             Length:         2012
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-858-332G-28 (1-6) x DROORFB (1-2012)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        538 CATAACCATAATCACAAC 521

RESULT 237
AR544899
LOCUS       AR544899                2091 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 30 from patent US 6747137.
ACCESSION   AR544899
VERSION     AR544899.1  GI:53938074
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2091)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 30 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .2091
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              476             Length:         2091
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x AR544899 (1-2091)
```

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1099 CACAACCACAACCACAAT 1116

RESULT 238
AB210685
LOCUS       AB210685                2127 bp    mRNA    linear   INV 08-JUL-2005
DEFINITION  Ciona intestinalis mRNA for Smad1/5, complete cds, clone:
            ciem812p06.
ACCESSION   AB210685
VERSION     AB210685.1  GI:70571162
KEYWORDS    .
SOURCE      Ciona intestinalis
  ORGANISM  Ciona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1
  AUTHORS   Satou,Y. and Satoh,N.
  TITLE     Genomewide surveys of developmentally relevant genes in Ciona
            intestinalis
  JOURNAL   Dev. Genes Evol. 213 (5-6), 211-212 (2003)
   PUBMED   12736827
REFERENCE   2
  AUTHORS   Imai,K.S., Hino,K., Yagi,K., Satoh,N. and Satou,Y.
  TITLE     Gene expression profiles of transcription factors and signaling
            molecules in the ascidian embryo: towards a comprehensive
            understanding of gene networks
  JOURNAL   Development 131 (16), 4047-4058 (2004)
   PUBMED   15269171
REFERENCE   3
  AUTHORS   Satou,Y.
  TITLE     Expressed genes in Ciona intestinalis
  JOURNAL   Unpublished
REFERENCE   4  (bases 1 to 2127)
  AUTHORS   Satou,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-APR-2005) Yutaka Satou, Kyoto University, Department
            of Zoology; Sakyo-ku, Kyoto, Kyoto, 606-8502, Japan
            (E-mail:yutaka@ascidian.zool.kyoto-u.ac.jp, Tel:81-75-753-4095,
            Fax:81-75-705-1113)
COMMENT     Please visit our home page
            URL:http://ghost.zool.kyoto-u.ac.jp/tfst.html.
FEATURES             Location/Qualifiers
     source          1. .2127
                     /organism="Ciona intestinalis"
                     /mol_type="mRNA"
                     /db_xref="taxon:7719"
                     /clone="ciem812p06"
     gene            1. .2127
                     /gene="Ci-Smad1/5"
     CDS             51. .1484
                     /gene="Ci-Smad1/5"
                     /codon_start=1
                     /product="Smad1/5"
                     /protein_id="BAE06690.1"
                     /db_xref="GI:70571163"
                     /translation="MAMTSLFGFTSPAVKRLLGWKQGDEEEKWAEKAVDSLVKKLKKK
                     KGSMEELERALSCPGQPSKCVTIPRSLDGRLQVSHRKGLPHVIYCRVWRWPDLQSHHE
                     LKPLEICSYPFDAKQKDVCINPYHYKRVESPVLPPVLVPRHSEFVPSHSMIHGNGGGH
                     NHNHNHMPPNANYPQNNSYNRPVPQLNHPTIPNQGPSSPYGGSQSTDVYSPNSFQPPQ
                     SPDGSVSGPNSPYSIPADTPPPAYMPPEGDKVVPGSEAMETNRTMISPNNRGIEVPVA
                     YSEPNHWCTIVYYELNNRVGEAFPATSTSIIVDGFTNPSSTSNRFCLGQLSNVNRNST
                     IENTRRHISKGVHLYYVGGEVYAECLSDSSIFVQSRNCNYHHGFHPTTVCKIPPSCSL
                     KIFNNQEFAALLSQSVNHGYEAVFELTKMCTIRMSFVKGWGAEYHRQDVTSTPCWIEI
                     HLNGPLQWLDKVLTQMGSPNNAISSVS"
ORIGIN

Alignment Scores:
Pred. No.:              484          Length:        2127
Score:                  42.00        Matches:       6
Percent Similarity:     100.00%      Conservative:  0
Best Local Similarity:  100.00%      Mismatches:    0
Query Match:            100.00%      Indels:        0
DB:                     2            Gaps:          0

US-09-858-332G-28 (1-6) x AB210685 (1-2127)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        528 CACAACCATAACCACAAC 545

RESULT 239
BC083830
LOCUS       BC083830                2152 bp    mRNA    linear   ROD 30-MAR-2005
DEFINITION  Rattus norvegicus deleted in polyposis 1-like 1 (predicted), mRNA
            (cDNA clone MGC:94922 IMAGE:7112359), complete cds.
ACCESSION   BC083830
VERSION     BC083830.1  GI:54035486
KEYWORDS    MGC.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 2152)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  CONSRTM   Mammalian Gene Collection Program Team
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2152)
  AUTHORS   .
  CONSRTM   NIH MGC Project
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) National Institutes of Health, Mammalian
            Gene Collection (MGC), Bethesda, MD 20892-2590, USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Howard Jacobs
            cDNA Library Preparation: Express Genomics
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 182 Row: h Column: 17
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Similarity but not
            identity to protein.
FEATURES             Location/Qualifiers
     source          1. .2152
```

```
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
                     /clone="MGC:94922 IMAGE:7112359"
                     /tissue_type="Testis, rat (Brown Norway)"
                     /clone_lib="NIH_MGC_237"
                     /lab_host="DH10B"
                     /note="Vector: pExpress1"
     gene            1. .2152
                     /gene="Dp111_predicted"
                     /db_xref="GeneID:362835"
                     /db_xref="RGD:1309508"
     CDS             28. .663
                     /gene="Dp111_predicted"
                     /codon_start=1
                     /product="deleted in polyposis 1-like 1 (predicted)"
                     /protein_id="AAH83830.1"
                     /db_xref="GI:54035487"
                     /db_xref="GeneID:362835"
                     /db_xref="RGD:1309508"
                     /translation="MDGLRQRFERFLEQKNVATDALGALEARTGVEKRYLAAGALTLL
                     GLYLLFGYGASLLCNVIGFVYPAYASVKAIESPNKEDDTVWLTYWVVYALFGLVEFFS
                     DLLLFWFPFYYAGKCAFLLFCMTPGPWNGALLLYHRVIRPLFLKHHVALDSAASQLSG
                     RALDIAAGITRDVLQALARGRTLVTPASASESPAALEPDPKSSQTTLLKHK"
ORIGIN
```

Alignment Scores:

Pred. No.:	489	Length:	2152
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

US-09-858-332G-28 (1-6) x BC083830 (1-2152)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        919 CACAACCACAACCACAAC 936
```

```
RESULT 240
AK069730/c
LOCUS       AK069730               2162 bp    mRNA    linear   PLN 24-JUL-2003
DEFINITION  Oryza sativa (japonica cultivar-group) cDNA clone:J023026G14, full
            insert sequence.
ACCESSION   AK069730
VERSION     AK069730.1  GI:32979754
KEYWORDS    FLI_CDNA; CAP trapper.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   The Rice Full-Length cDNA Consortium, National Institute of
            Agrobiological Sciences Rice Full-Length cDNA Project Team:,
            Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K.,
            Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I.,
            Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C.,
            Ohtsuki,K., Shishiki,T., Foundation of Advancement of International
            Science Genome Sequencing & Analysis Group:, Otomo,Y., Murakami,K.,
            Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y.,
            Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M.,
            Narikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Niikura,J.,
            Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Miura,J.,
            Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN:,
            Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuda,S.,
            Hara,A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M.,
            Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Osato,N., Ota,Y.,
            Saito,R., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
            Yoshino,M. and Hayashizaki,Y.
  TITLE     Collection, mapping, and annotation of over 28,000 cDNA clones from
            japonica rice
  JOURNAL   Science 301 (5631), 376-379 (2003)
  PUBMED    12869764
REFERENCE   2  (bases 1 to 2162)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K.,
            Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiraoka,T.,
            Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K.,
            Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I.,
            Kanagawa,S., Katoh,H., Kawagashira,N., Kawai,J., Kawamata,M.,
            Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M.,
            Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Kurosaki,T., Kusumegi,T., Li,C., Lu,M.,
            Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A.,
            Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakamura,M.,
            Namiki,T., Narikawa,R., Niikura,J., Nishi,K., Nomura,K.,
            Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M., Ooka,H.,
            Osato,N., Ota,Y., Otomo,Y., Ryu,R., Saitoh,H., Sakai,C., Sakai,K.,
            Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Shibata,K.,
            Shinagawa,A., Shiraki,T., Shishiki,T., Sogabe,Y., Sugano,S.,
            Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y.,
            Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A.,
            Toya,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,W.,
            Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and
            Yoshimura,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of
            Agrobiological Sciences, Department of Molecular Genetics, Head of
            Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki
            305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp,
            Tel:81-29-838-7007, Fax:81-29-838-7007)
COMMENT     This clone is one of the 28K full-length cDNA clones from japonica
            rice.
            URL : http://cdna01.dna.affrc.go.jp/cDNA/
            NIAS Rice Full-Length cDNA Project Team: Kikuchi,S., Satoh,K.,
            Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J.,
            Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T.,
            Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T. and
            Yamamoto,M.
            FAIS Genome Sequencing & Analysis Group: Otomo,Y., Iida,Y.,
            Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M., Kobayashi,M.,
            Kodama,T., Kurosaki,T., Kusumegi,T., Lu,M., Masuda,H., Miura,J.,
            Mizuno,K., Narikawa,R., Niikura,J., Oka,M., Ryu,R., Sugano,S.,
            Sugiyama,A., Suzuki,Y., Tsunoda,Y., Ueda,M., Xie,Q., Yokomizo,S.,
            Yoshimura,A., Matsubara,K. and Murakami,K.
            Genome Exploration Research Group in Riken Genomic Sciences Center
            and Genome Science Laboratory in Riken: Adachi,J., Aizawa,K.,
            Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hanagaki,T.,
            Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K.,
            Hiraoka,T., Hori,F., Iida,J., Imamura,K., Imotani,K., Ishii,Y.,
            Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J.,
            Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Osato,N.,
            Ota,Y., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H.,
            Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F.,
            Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K.,
            Yasunishi,A. and Hayashizaki,Y.
FEATURES             Location/Qualifiers
     source          1. .2162
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="J023026G14"
ORIGIN
```

Alignment Scores:

Pred. No.:	492	Length:	2162
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0

```
DB:                    15            Gaps:          0

US-09-858-332G-28 (1-6) x AK069730 (1-2162)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        325 CACAACCACAACCACAAC 308

RESULT 241
SCYLL010C/c
LOCUS       SCYLL010C               2212 bp    DNA     linear   PLN 11-AUG-1997
DEFINITION  S.cerevisiae chromosome XII reading frame ORF YLL010c.
ACCESSION   Z73115 Y13138
VERSION     Z73115.1  GI:1360174
KEYWORDS    .
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 2212)
  AUTHORS   Miosga,T. and Zimmermann,F.K.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2212)
  AUTHORS   MIPS.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAY-1996) Data collected by MIPS on behalf of the
            European yeast chromosome XII sequencing project. MIPS at the
            Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
            Martinsried, FRG; E-mail: Mewes@mips.embnet.org
FEATURES             Location/Qualifiers
     source          1. .2212
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4932"
                     /chromosome="XII"
     CDS             complement(338. .1621)
                     /note="unnamed protein product; ORF YLL010c"
                     /codon_start=1
                     /protein_id="CAA97454.1"
                     /db_xref="GI:1360175"
                     /db_xref="GOA:Q07800"
                     /db_xref="SGD:S0003933"
                     /db_xref="UniProt/TrEMBL:Q07800"
                     /translation="MGFISSILCCSSETTQSNSNSAYRQQQSSSLNKNRSVKHSNTKS
                     RTRGVHQTNSPPSKTNSAATFSSTERSTGKSGISTNDNEKKKPSSPTAAVTATTTNNM
                     TKVEKRISKDDLYEEKYEVDEDEEIDDEDNRRSRGIVQEKGDAVKDTSRQKKQQQQQQ
                     QQSQPQPQPQSQSQSQSQSQSQQRGPTVQVSSDHLIQDMNLSRVSSSSQASETSNDAD
                     DEDDEDEEYIDLTLLQQGQYHAPGYNTLLPPQDESTKGKKCLILDLDETLVHSSFKYL
                     RSADFVLSVEIDDQVHNVYVIKRPGVEEFLERVGKLFEVVVFTASVSRYGDPLLDILD
                     TDKVIHHRLFREACYNYEGNYIKNLSQIGRPLSDIIILDNSPASYIFHPQHAIPISSW
                     FSDTHDNELLDIIPLLEDLSVKTSLDVGKILDVTI"
ORIGIN

Alignment Scores:
Pred. No.:              503          Length:        2212
Score:                  42.00        Matches:       6
Percent Similarity:     100.00%      Conservative:  0
Best Local Similarity:  100.00%      Mismatches:    0
Query Match:            100.00%      Indels:        0
DB:                     15           Gaps:          0

US-09-858-332G-28 (1-6) x SCYLL010C (1-2212)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1134 CACAACCACAACCACAAC 1117

RESULT 242
CPTKT7/c
LOCUS       CPTKT7                  2232 bp    mRNA    linear   PLN 18-APR-2005
DEFINITION  C.plantagineum tkt7 gene for transketolase.
ACCESSION   Z46648
VERSION     Z46648.1  GI:664902
KEYWORDS    tkt7 gene; transketolase.
SOURCE      Craterostigma plantagineum
  ORGANISM  Craterostigma plantagineum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Lamiales; Linderniaceae; Craterostigma.
REFERENCE   1  (bases 1 to 2232)
  AUTHORS   Bernacchia,G., Schwall,G., Lottspeich,F., Salamini,F. and
            Bartels,D.
  TITLE     The transketolase gene family of the resurrection plant
            Craterostigma plantagineum: differential expression during the
            rehydration phase
  JOURNAL   EMBO J. 14 (3), 610-618 (1995)
   PUBMED   7859749
REFERENCE   2  (bases 1 to 2232)
  AUTHORS   Bartels,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-NOV-1994) Bartels D., Max-Planck-Institut fuer
            Zuechtungsforschung, Pflanzenzuechtung und Ertragsphysiologie,
            Carl-von-Linne-Weg 10, Cologne, NRW, Germany, 50829
FEATURES             Location/Qualifiers
     source          1. .2232
                     /organism="Craterostigma plantagineum"
                     /mol_type="mRNA"
                     /db_xref="taxon:4153"
                     /clone="Cpl-tkt7"
     gene            1. .2232
                     /gene="Cpl-tkt7"
     CDS             29. .2059
                     /gene="Cpl-tkt7"
                     /EC_number="2.2.1.1"
                     /citation=[1]
                     /codon_start=1
                     /product="transketolase"
                     /protein_id="CAA86609.1"
                     /db_xref="GI:664903"
                     /db_xref="GOA:Q42677"
                     /db_xref="InterPro:IPR005474"
                     /db_xref="InterPro:IPR005475"
                     /db_xref="InterPro:IPR005476"
                     /db_xref="InterPro:IPR005478"
                     /db_xref="InterPro:IPR009014"
                     /db_xref="UniProt/Swiss-Prot:Q42677"
                     /translation="MAPKTTLIAEPELVSKSVNTIRFLAIDAVEKAKSGHPGMPMGCA
                     PMGHVLYDEFMRFNPKNPYWFNRDRFVLSAGHGCMLQYALLHLSGYDSVKEEDLKSLR
                     QWGSRTPAHPENFETPGVEVTTGPLGQGIASAVGLAVAEKHLAARYNKPGFEIVDHYT
                     YVILGDGCQMEGVSNEACSLAAHWGLGKLIALYDDNHITIDGDTDVAFTEDVDKRFDA
                     LGWHVIWVKNGNDGCDEIRAAIEEAKSVKDRPTMIKVTTTIGYGAPSKANTYGVHGNA
                     LGPKEAEATRKNLGWPYEPFHVPDDVKKHWSRHIAEGAALESAWNAKFAEFQKKFPEE
                     AADLKSIITGELPTNWESIFPTYTPENPGLPTRTLSHQILNGLGDVLPGLLGGSADLT
                     LSNMAFLKNSGDFQKKSPGERNVKFGAREHAMGSICNGLALHSPGLLPYCATYFVFTD
                     YMRAAMRISALSKARVLYIMTHDSIGLGEDGPTHQPVEHLASFRAMPNILTLRPADGN
                     ETAGAYRAAVQNGERPSILVLARQKLPQLPGTSIEGVSKGGYVISDNSRGGNSKPDVI
                     LIGTGSELEIAAARAGDELRKEGKKVRVVSLVCWELFAEQSEKYRETVLPSGVTARVSV
                     EAGSTFGWERFIGPKGKAVGIDRFGASAPAERLFKEFGITVEAVVAAAKEIC"
ORIGIN

Alignment Scores:
Pred. No.:              507          Length:        2232
Score:                  42.00        Matches:       6
Percent Similarity:     100.00%      Conservative:  0
Best Local Similarity:  100.00%      Mismatches:    0
Query Match:            100.00%      Indels:        0
DB:                     15           Gaps:          0

US-09-858-332G-28 (1-6) x CPTKT7 (1-2232)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       2153 CACAATCACAATCACAAT 2136

RESULT 243
AX461526/c
```

```
US-09-858-332G-28 (1-6) x AK113909 (1-2362)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        525 CACAACCATAACCACAAC 542

RESULT 246
AF020407
LOCUS       AF020407               2396 bp    DNA     linear   INV 11-MAY-1998
DEFINITION  Dictyostelium discoideum WimA (wimA) gene, partial cds.
ACCESSION   AF020407 AF012944
VERSION     AF020407.1  GI:2425142
KEYWORDS    .
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 2396)
  AUTHORS   Loomis,W.F. and Iranfar,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-AUG-1997) Biology, UCSD, 9500 Gilman Drive, La Jolla,
            CA 92093, USA
COMMENT     On May 11, 1998 this sequence version replaced gi:2393950.
FEATURES             Location/Qualifiers
     source          1. .2396
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
     gene            702. .>2396
                     /gene="wimA"
                     /note="disruption results in morphological defect;
                     developmental gene"
     CDS             702. .>2396
                     /gene="wimA"
                     /codon_start=1
                     /product="WimA"
                     /protein_id="AAC16025.1"
                     /db_xref="GI:2425143"
                     /translation="MENQDSSLFFGEPDPQIDFFDSLSSGQEQQEQQQEQQQEQQQEQ
                     QEQTQPQTQQEQPTSHFEDGESLFESLGQQDSNSTNTTTNIESPPHQQDDQSFFDSFN
                     SSDSGVGITNEQPKTEQPTQPFSNLIQPNEVIELQSDVIQHQNQVIEHQNSIIEQHQR
                     HQHQFSSDSSNIGIVDSLIENPILVSDNVVNSLSDSIVGKDDDETSPNNNNIDNIDNI
                     DNIDNIDDNNNNTNNNNNSGNQENENLSTSSSNEHTPAQSEPIVTTTEGSNNNENNI
                     INNNENISRDDDNSNSNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNSGEKEK
                     EISSFFSSDDNVNNSEVLFFNSTNSDTNSSANSIETVSVTSAIAAEEKDASSFFNNIS
                     ASQELQPQPQPQPQPQPQPQPQPQPQPQPQQQTQQQQSQTQQSPITNKPPVVEKDISSFF
                     ATPEMESQNDPSGNIFQSFNISSNTPKMVSQVPQPQPQQQTQPPPPQQQQQPQQPLQT
                     PTKSQPQQTEKDVSSFFGGGSSSFKQPQQPERDVSSFFGGGSTTNEVSSLFGSMTID"
ORIGIN

Alignment Scores:
Pred. No.:              543           Length:         2396
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     2             Gaps:           0

US-09-858-332G-28 (1-6) x AF020407 (1-2396)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1897 CACAACCACAACCACAAC 1914

RESULT 247
AR672184
LOCUS       AR672184               2396 bp    DNA     linear   PAT 13-JUN-2005
DEFINITION  Sequence 1685 from patent US 6902887.
ACCESSION   AR672184
VERSION     AR672184.1  GI:67611980
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2396)
  AUTHORS   Berka,R.M., Rey,M.W., Shuster,J.R., Kauppinen,S., Clausen,I.G. and
            Olsen,P.B.
  TITLE     Methods for monitoring multiple gene expression
  JOURNAL   Patent: US 6902887-A 1685 07-JUN-2005;
            Novozymes Biotech, Inc. and Novozymes A/S; Davis, CA
FEATURES             Location/Qualifiers
     source          1. .2396
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              543           Length:         2396
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     6             Gaps:           0

US-09-858-332G-28 (1-6) x AR672184 (1-2396)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        190 CATAACCATAACCATAAT 207

RESULT 248
AF036170
LOCUS       AF036170               2400 bp    mRNA    linear   INV 20-DEC-1997
DEFINITION  Dictyostelium discoideum homeobox-containing protein Wariai mRNA,
            partial cds.
ACCESSION   AF036170
VERSION     AF036170.1  GI:2707267
KEYWORDS    .
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 2400)
  AUTHORS   Han,Z. and Firtel,R.A.
  TITLE     The homeobox-containing gene Wariai regulates anterior-posterior
            patterning and cell-type homeostasis in Dictyostelium
  JOURNAL   Development (1998) In press
REFERENCE   2  (bases 1 to 2400)
  AUTHORS   Han,Z. and Firtel,R.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-NOV-1997) Biology, UCSD, 9500 Gilman Drive, La Jolla,
            CA 92093, USA
FEATURES             Location/Qualifiers
     source          1. .2400
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="KAx3"
                     /db_xref="taxon:44689"
     gene            1. .>2400
                     /gene="wariai"
     CDS             1. .>2400
                     /gene="wariai"
                     /note="stalk-cell specific; contains ankyrin repeats"
                     /codon_start=1
                     /product="homeobox-containing protein Wariai"
                     /protein_id="AAB92245.1"
                     /db_xref="GI:2707268"
                     /translation="MASIVMKSQHSLGYPNIGNINRSDYDSYEQQYNNPTGSKQYNNN
                     NNNNTNTNEINNGTNTNLNPNNMYGMYNNNNNNNNNNNNNNNNNNNNSNNNNNNNN
                     NNNINNNNNNNNNNNNNNNNNNQHLSQSQQLSPTPYSSNSFSKLLSRSTNDLMLDQDD
                     PSKKKRKRTSPDQLKLLEKIFMAHQHPNLNLRSQLAVELHMTARSVQIWFQNRRAKAR
                     NMEFKPQLSHGGSDLIYNALGSQNGMNSMGGGGGNGGGNGGGGMNGINNILNGNHNRN
                     LNKYIPHGGNSINGNMGGGGGGGGGSHNHNHNHNHNHNHNHNHNQPLSNGDCGEKFS
                     VASAWNKILLHPNNIDFLIRYNPDDPNSIDVNARDSKGLSLLFTAAFLGYEYQVRRLI
                     ESGANPNIKDNQGDTPLIAASVLGNQPIVELLLEHRADPNLVNDEGVSPLFSACKGGH
                     LQIXSSLLDHDREVSVKTKINGETPLHIASLKGFEKICKLLIETEAKASVIDSNNRTP
```

```
LOCUS       AX461526               2270 bp    DNA     linear   PAT 08-JUL-2002
DEFINITION  Sequence 455 from Patent WO0198480.
ACCESSION   AX461526
VERSION     AX461526.1  GI:21726737
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Budworth,P., Brown,D., Chang,H.S., Zhu,T., Han,B., Wang,X. and
            Cooper,B.
  TITLE     Promoters for regulation of plant gene expression
  JOURNAL   Patent: WO 0198480-A 455 27-DEC-2001;
            Syngenta Participations AG (CH)
FEATURES             Location/Qualifiers
     source          1. .2270
                     /organism="Arabidopsis thaliana"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:3702"
ORIGIN

Alignment Scores:
Pred. No.:                515           Length:         2270
Score:                    42.00         Matches:        6
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              100.00%       Indels:         0
DB:                       6             Gaps:           0

US-09-858-332G-28 (1-6) x AX461526 (1-2270)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        604 CACAATCACAATCACAAT 587

RESULT 244
AF191551/c
LOCUS       AF191551               2347 bp    mRNA    linear   ROD 03-NOV-1999
DEFINITION  Mus musculus putative Ca influx channel beta variant (TRPC1A-Beta)
            mRNA, complete cds.
ACCESSION   AF191551
VERSION     AF191551.1  GI:6179957
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2347)
  AUTHORS   Rae,J.L.
  TITLE     Ion channels in lens epithelia
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2347)
  AUTHORS   Rae,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-1999) Physiology and Biophysics, Mayo Foundation,
            200 1st Street SW, Rochester, MN 55905, USA
FEATURES             Location/Qualifiers
     source          1. .2347
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /cell_line="alphaTN4"
                     /tissue_type="lens epithelium"
     gene            1. .2347
                     /gene="TRPC1A-Beta"
     CDS             20. .2347
                     /gene="TRPC1A-Beta"
                     /codon_start=1
                     /product="putative Ca influx channel beta variant"
                     /protein_id="AAF05725.1"
                     /db_xref="GI:6179958"
                     /translation="MGAPPPSPGLPPSWAAMMAALYPSTDLSGVSSSSLPSSPSSSSP
                     NEVMALKDVREVKEENTLNEKLFLLACDKGDYYMVKKILEENSSGDLNINCVDVLGRN
                     AVTITIENESLDILQLLLDYGCQKLMERIQNPEYSTTMDVAPVILAAHRNNYEILTML
                     LKQDVSLPKPHAVGCECTLCSAKNKKDSLRHSRFRLDIYRCLASPALIMLTEEDPILR
                     AFELSADLKELSLVEVEFRNDYEELARQCKMFAKDLLAQARNSRELEVILNHTSSDEP
                     LDKRGLLEERMNLSRLKLAIKYNQKEFVSQSNCQQFLNTVWFGQMSGYRRKPTCKKIM
                     TVLTVGIFWPVLSLCYLIAPKSQFGRIIHTPFMKFIIHGASYFTFLLLLNLYSLVYNE
                     DKKNTMGPALERIDYLLILWIIGMIWSDIKRLWYEGLEDFLEESRNQLSFVMNSLYLA
                     TFALKVVAHNKFHDFADRKDWDAFHPTLVAEGLFAFANVLSYLRLFFMYTTSSILGPL
                     QISMGQMLQDFGKFLGMFLLVLFSFTIGLTQLYDKGYTSKEQKDCVGIFCEQQSNDTF
                     HSFIGTCFALFWYIFSLAHVAIFVTRFSYGEELQSFVGAVIVGTYNVVVVIVLTKLLV
                     AMLHKSFQLIANHEDKEWKFARAKLWLSYFDDKCTLPPPFNIIPSPKTICYMISSLSK
                     WICSHTSKGKVKRQNSLKEWRNLKQKRDENYQKVMCCLVHRYLTSMRQKMQSTDQATV
                     ENLNELRQDLSKFRNEIRDLLGFRTSKYAMFYPRN"
ORIGIN

Alignment Scores:
Pred. No.:                532           Length:         2347
Score:                    42.00         Matches:        6
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              100.00%       Indels:         0
DB:                       9             Gaps:           0

US-09-858-332G-28 (1-6) x AF191551 (1-2347)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1873 CACAATCACAACCACAAC 1856

RESULT 245
AK113909
LOCUS       AK113909               2362 bp    mRNA    linear   INV 30-NOV-2002
DEFINITION  Ciona intestinalis cDNA, clone:cicl016e21, full insert sequence.
ACCESSION   AK113909
VERSION     AK113909.1  GI:23584999
KEYWORDS    FLI_CDNA.
SOURCE      Ciona intestinalis
  ORGANISM  Ciona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1
  AUTHORS   Satou,Y., Yamada,L., Mochizuki,Y., Takatori,N., Kawashima,T.,
            Sasaki,A., Hamaguchi,M., Awazu,S., Yagi,K., Sasakura,Y.,
            Nakayama,A., Ishikawa,H., Inaba,K. and Satoh,N.
  TITLE     A cDNA resource from the basal chordate Ciona intestinalis
  JOURNAL   Genesis 33 (4), 153-154 (2002)
  PUBMED    12203911
REFERENCE   2  (bases 1 to 2362)
  AUTHORS   Satou,Y. and Satoh,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-OCT-2002) Nori Satoh, Kyoto University, Department of
            Zoology; Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            (E-mail:satoh@ascidian.zool.kyoto-u.ac.jp, Tel:81-75-753-4095,
            Fax:81-75-705-1113)
COMMENT     Ciona intestinalis cDNA Project (URL:
            http://ghost.zool.kyoto-u.ac.jp/indexr1.html).
FEATURES             Location/Qualifiers
     source          1. .2362
                     /organism="Ciona intestinalis"
                     /mol_type="mRNA"
                     /db_xref="taxon:7719"
                     /clone="cicl016e21"
ORIGIN

Alignment Scores:
Pred. No.:                535           Length:         2362
Score:                    42.00         Matches:        6
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              100.00%       Indels:         0
DB:                       2             Gaps:           0
```

LHHACIMGYFSIAKLLICNGADMNAIDIDGHTPLHTSSLMGQYLITRLLLENGADPNI
QDSEGYTPIHYAVRESRIETVKFLIKFNSKLNIKTKNGQNLIHLSVQFASLMMGQMIF
ESKGCEIAADDSDDQGYTPLYLAAKAGKTNFVKYLLSKGRSKKIRLEKLIQENQDKEI
IQMLESTVTKSSNNNNSNSNINNINNINNINNINSQPNTNSDNNNNNNNNNFNENYSN
GNNEQSQPPGNKFEEDDEDDFYDRVYKKSYTNRIISNSFSYQQKLNSGNGISVNSGNL
ED"

ORIGIN

Alignment Scores:

Pred. No.:	544	Length:	2400
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-28 (1-6) x AF036170 (1-2400)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        907 CATAATCATAATCATAAT 924
```

RESULT 249
AK070717
LOCUS AK070717 2475 bp mRNA linear PLN 24-JUL-2003
DEFINITION Oryza sativa (japonica cultivar-group) cDNA clone:J023058L06, full insert sequence.
ACCESSION AK070717
VERSION AK070717.1 GI:32980741
KEYWORDS FLI_CDNA; CAP trapper.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1
AUTHORS The Rice Full-Length cDNA Consortium, National Institute of Agrobiological Sciences Rice Full-Length cDNA Project Team:, Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group:, Otomo,Y., Murakami,K., Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y., Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M., Narikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Niikura,J., Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Miura,J., Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN:, Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Osato,N., Ota,Y., Saito,R., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Yoshino,M. and Hayashizaki,Y.
TITLE Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice
JOURNAL Science 301 (5631), 376-379 (2003)
PUBMED 12869764
REFERENCE 2 (bases 1 to 2475)
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K., Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K., Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawagashira,N., Kawai,J., Kawamata,M., Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M., Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Kurosaki,T., Kusumegi,T., Li,C., Lu,M., Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A., Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakamura,M., Namiki,T., Narikawa,R., Niikura,J., Nishi,K., Nomura,K., Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M., Ooka,H., Osato,N., Ota,Y., Otomo,Y., Ryu,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Shibata,K., Shinagawa,A., Shiraki,T., Shishiki,T., Sogabe,Y., Sugano,S., Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,W., Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and Yoshimura,A.
TITLE Direct Submission
JOURNAL Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)
COMMENT This clone is one of the 28K full-length cDNA clones from japonica rice.
URL : http://cdna01.dna.affrc.go.jp/cDNA/
NIAS Rice Full-Length cDNA Project Team: Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T. and Yamamoto,M.
FAIS Genome Sequencing & Analysis Group: Otomo,Y., Iida,Y., Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M., Kobayashi,M., Kodama,T., Kurosaki,T., Kusumegi,T., Lu,M., Masuda,H., Miura,J., Mizuno,K., Narikawa,R., Niikura,J., Oka,M., Ryu,R., Sugano,S., Sugiyama,A., Suzuki,Y., Tsunoda,Y., Ueda,M., Xie,Q., Yokomizo,S., Yoshimura,A., Matsubara,K. and Murakami,K.
Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Iida,J., Imamura,K., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J., Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Osato,N., Ota,Y., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K., Yasunishi,A. and Hayashizaki,Y.

FEATURES Location/Qualifiers
source 1. .2475
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="J023058L06"

ORIGIN

Alignment Scores:

Pred. No.:	560	Length:	2475
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	15	Gaps:	0

US-09-858-332G-28 (1-6) x AK070717 (1-2475)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        379 CACAACCACAACCACAAC 396
```

RESULT 250
BT001630
LOCUS BT001630 2506 bp mRNA linear INV 15-NOV-2002
DEFINITION Drosophila melanogaster RE38276 full insert cDNA.
ACCESSION BT001630
VERSION BT001630.1 GI:25012569
KEYWORDS FLI_CDNA.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

```
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 2506)
  AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J.,
            Champe,M., Chavez,C., Dorsett,V., Dresnek,D., Farfan,D., Frise,E.,
            George,R., Gonzalez,M., Guarin,H., Kronmiller,B., Li,P., Liao,G.,
            Miranda,A., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
            Patel,S., Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Rubin,G.M.
            and Celniker,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2002) Berkeley Drosophila Genome Project,
            Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, USA
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
             Lawrence Berkeley National Laboratory
            Berkeley, CA 94720
             This clone was sequenced as part of a high-throughput process to
            sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
            Science 2000). The sequence has been subjected to integrity checks
            for sequence accuracy, presence of a polyA tail and contiguity
            within 100 kb in the genome. Thus we believe the sequence to
            reflect accurately this particular cDNA clone. However, there are
            artifacts associated with the generation of cDNA clones that may
            have not been detected in our initial analyses such as internal
            priming, priming from contaminating genomic DNA, retained introns
            due to reverse transcription of unspliced precursor RNAs, and
            reverse transcriptase errors that result in single base changes.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our Web site
            (http://fruitfly.berkeley.edu) or send email to
            cdna@fruitfly.berkeley.edu.
FEATURES                Location/Qualifiers
     source             1. .2506
                        /organism="Drosophila melanogaster"
                        /mol_type="mRNA"
                        /strain="y; cn bw sp"
                        /db_xref="taxon:7227"
     gene               1. .2506
                        /gene="CG5169"
                        /db_xref="FLYBASE:FBgn0038477"
     CDS                301. .2229
                        /gene="CG5169"
                        /note="Longest ORF"
                        /codon_start=1
                        /product="RE38276p"
                        /protein_id="AAN71385.1"
                        /db_xref="GI:25012570"
                        /db_xref="FLYBASE:FBgn0038477"
                        /translation="MSWTEKVDPELIFTKQERIGKGSFGEVFKGIDNRTQQVVAIKII
                        DLEEAEDEIDDIQQEIMVLSQCDSPYVTKYYGSFLKGTKLWIIMEYLGGGSALDLMKA
                        GSFEEMHIGIILREVLKGLDYLHSERKLHRDIKAANVLLSEQGDVKLADFGVAGQLTN
                        TTSKRNTFVGTPFWMAPEVIKQSQYDAKADIWSLGITAIELAKGEPPNSELHPMRVLF
                        LIPKNNPPQLTGSYTKSFKDFVEACLNKDPENRPTAKELLKYPFIKKAKKNAYLIDLI
                        DRFKKWKVSKGDESETESENSDSDSDAKQGGSSQDEPWIMTVKGLHINSAPRAGVNSF
                        QLQEHELTMQKLMQTTHSPAGGGGVQASPAANALSSSVSAVAASASSVSVITANEGVS
                        SASALPPQNSHVHNHNHNNINNNNIRNLSNKHATTTMLNAAPPPSSVTSSSSASELQQ
                        KRSSSKPELRQSRSAAPLDQVEDRRASLGLPLQSSVPTPSTAEQQQQRRSSQRESFHA
                        NNAALVASCSSSSNNHSHEEHVATNHASHVKHSGAGQRLVQTNSACLNNFLFPVLSDI
                        QREYSGGGSAGSGGSGGRNIGVGSDIGDLKSVFELAERSSPGVSDLFMKELIHMLVPG
                        YSETRINTIMDRAIRNRK"
ORIGIN

Alignment Scores:
Pred. No.:                567             Length:         2506
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       2               Gaps:           0

US-09-858-332G-28 (1-6) x BT001630 (1-2506)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1513 CACAACCACAACCACAAC 1530

RESULT 251
AF339902S2
LOCUS       AF339902S2              2541 bp    DNA     linear   INV 12-NOV-2001
DEFINITION  Dictyostelium discoideum phosphatidylinositol phosphate kinase 6
            (PIPkinA) gene, exon 2 and partial cds.
ACCESSION   AF339903
VERSION     AF339903.1  GI:16903268
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 2541)
  AUTHORS   Guo,K., Nichol,R., Skehel,P., Dormann,D., Weijer,C.J.,
            Williams,J.G. and Pears,C.
  TITLE     A Dictyostelium nuclear phosphatidylinositol phosphate kinase
            required for developmental gene expression
  JOURNAL   EMBO J. 20 (21), 6017-6027 (2001)
   PUBMED   11689442
REFERENCE   2  (bases 1 to 2541)
  AUTHORS   Nichol,R., Guo,K., Skehel,P., Williams,J.G. and Pears,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JAN-2001) Biochemistry Department, Oxford University,
            South Parks Road, Oxford, Oxfordshire OX1 3QU, England
FEATURES                Location/Qualifiers
     source             1. .2541
                        /organism="Dictyostelium discoideum"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:44689"
     gene               order(AF339902.1:<1. .2226,1. .>2541)
                        /gene="PIPkinA"
     mRNA               join(AF339902.1:<1. .2226,1. .>2541)
                        /gene="PIPkinA"
                        /product="phosphatidylinositol phosphate kinase 6"
     CDS                join(AF339902.1:1. .2226,1. .>2541)
                        /gene="PIPkinA"
                        /codon_start=1
                        /product="phosphatidylinositol phosphate kinase 6"
                        /protein_id="AAL30450.1"
                        /db_xref="GI:16903269"
                        /translation="MTICQPIPCGGSYFWNIVEPYKNDEYVFVMQSYNGPFRWKSEEY
                        RIVLMCGPNSNGYIIATSTLFSEIFSNWEIITTIYKEKIEKTGLSHFSDTFLIQLYKK
                        LAKKLSQVNSPIFTEEEQNIQPPPQPQLPPQQPPPPPQIQESQPIQEQIEQIILPTTL
                        EFPKSRDNEEIKIIKKRKIVALFKTLYRVDAMTIIERDYSSSSQWIEWVDESTGMRPI
                        HVAVERMNVQLVKYLIERKADVNIKDNQGWSPLHFSSFAGSLDICQILLDQGNAAVLI
                        ISKDGTLPLHYLIRHCYSINVPSSSSNSNNNKDNNNNNYKDNNKFNSNLSINYNNGGD
                        CNNLTLKQKQESNNKLFSILSLLLSKGTPINAKTIRGETALHRACYYGSAQSVKFLHQ
                        NGADVNVQNSRGETPLYFAVVSRQREIVKLLIEYASDVNIGGERSALKAADKTNQNEI
                        YYFLAGFSDEKSVSDSKRDYYNNNDENDSNNNNNNYNDENDNIHRHNQNNQGNNQNHS
                        MDCCSKIENGQECSCGDLSNTAHPFYPHVFVFIDFPAGTTHCSYCKYLLWGIRKQGFQ
                        CEVCSYIVHSRCKRQATLTNTCGIPDSKETISSSVVNDFLTKSRDQNINNNNEEEGVE
                        VVEEDHHNNINNNINNNNKRSRQRKQTIRQPLHKQNINRKRLESLYNHFITLDKEKK
                        GSILKKDFEKCLGPIINSSESLSNALFLGFNPKKHDKMSYVEFLTGVSVLQNSTFDKE
                        IQFSFKMLAGEKGYITVEEFLSILESIYSSLTNLTIVTCNPQAFLKRLFPEFSFSYQR
                        KQKLLLQQQQQLQQQQQQQQQQQQQKSSWSSTSPSPSSTTNSLRNSLRLSRAIDNNNN
                        HNNHNHNHNNSNINNEDDLNLDSDSDSDSFPTPSTSPLLFKIGIKNKQQQQQQQQQQQ
                        QQQQQQQQQQQQQQQQQQEQQQQQQQENQQIIKDENETVENQSQEIKQKEDENQKEVE
                        NQKQDLKNQKEEKEEEEEIKEQRKFETKNDNFVQVNTKKSVRGSIFFAPLDSHEKINN
                        ISSIGNHRSLIRNSSSSVLDNNSNNSNNNNNNKNNNESTATPNTTNSTTPITSATTST
                        SSSPSSSYKSQSLPDLPSHIPKTPTKSSIINNNNNTSKTITTKSSSSFENGLNNSGID
                        NNYISKKEKIENDKVFQLNGRIYFKEFKQALSDNLYFVKSLGLVNHYENPLIRETEGN
                        DGLLVNHSSNWVTSQGKDVSIGHINWELIQYIMIGIRRSAGEAIVLTNRATLKPKDFE
                        MVVEFKYDGWTFKDHYPLAFKKIRERLEIDPKMFMFSLGPERVFGNLLLGNLSVLSEM
                        NSSGKSGSVFFRSTEGDYLIKTIPTHEESILKAVLPTYVQHLQKYPNSLLIKILGCYT
                        LQIKGKAEMKFLVMNNLFFTPLPLSEKYDLKGSVINRKVDKNDLLLPDIALKDQEFHR
                        ILDIGPEFKAPLLEQIEHDTMFLESHNICDYSLLVGIHTIDENSPLALSDDDDPDLSN
                        VGGVKRDTWKVLEEEFFKKTSGKISLFQKNFGGILSKNKKEVYFIAIIDTFTAWDWWK
                        KSERALKFLGNDLDKISAVNPTDYRKRFQHYVSKIVQ"
     exon               1. .2541
```

```
                     /gene="PIPkinA"
                     /number=2
ORIGIN

Alignment Scores:
Pred. No.:                574             Length:         2541
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       2               Gaps:           0

US-09-858-332G-28 (1-6) x AF339902S2 (1-2541)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        352 CATAACCATAATCATAAT 369

RESULT 252
CQ586446
LOCUS       CQ586446                2555 bp    DNA     linear   PAT 02-FEB-2004
DEFINITION  Sequence 14204 from Patent WO0171042.
ACCESSION   CQ586446
VERSION     CQ586446.1  GI:41646277
KEYWORDS    .
SOURCE      Drosophila sp.
  ORGANISM  Drosophila sp.
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1
  AUTHORS   Venter,J.C., Adams,M., Li,P.W. and Myers,E.W.
  TITLE     Detection kits, such as nucleic acid arrays, for detecting the
            expression of 10,000 or more Drosophila genes and uses thereof
  JOURNAL   Patent: WO 0171042-A 14204 27-SEP-2001;
            PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1. .2555
                     /organism="Drosophila sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:7242"
ORIGIN

Alignment Scores:
Pred. No.:                577             Length:         2555
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x CQ586446 (1-2555)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        734 CATAATCACAATCACAAT 751

RESULT 253
DMRNAH15
LOCUS       DMRNAH15                2563 bp    mRNA    linear   INV 18-APR-2005
DEFINITION  D.melanogaster mRNA for H15 protein.
ACCESSION   X98766
VERSION     X98766.1  GI:1524007
KEYWORDS    .
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1
  AUTHORS   Brook,W.J. and Cohen,S.M.
  TITLE     Antagonistic interactions between wingless and spp organioze dorsal
            ventral pattern in the drospohila leg
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2563)
  AUTHORS   Brook,W.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUN-1996) W.J. Brook, European Molecular Biology
            Laboratory, Developmental Biology Programme, Meyerhofstrasse 1,
            Heidelberg, 69117, FRG
FEATURES             Location/Qualifiers
     source          1. .2563
                     /organism="Drosophila melanogaster"
                     /mol_type="mRNA"
                     /db_xref="taxon:7227"
                     /tissue_type="imaginal disc"
                     /dev_stage="larvae"
     CDS             886. .2520
                     /codon_start=1
                     /product="H15 protein"
                     /protein_id="CAA67304.1"
                     /db_xref="GI:1524008"
                     /db_xref="FLYBASE:FBgn0016660"
                     /db_xref="GOA:Q94890"
                     /db_xref="InterPro:IPR001699"
                     /db_xref="InterPro:IPR002070"
                     /db_xref="InterPro:IPR008967"
                     /db_xref="UniProt/Swiss-Prot:Q94890"
                     /translation="MRTMMLTLMWSTAVTRKCRRQLRQRRGGNNSSGCSGSIAGAATS
                     ASGIACCTATTAAAATTKAADTSPRNNWQATTPTSQSSQQQHQQQQQLRQQQHQQQIK
                     QPARAFAAAVGAAATPSPPPPPPSQSPEELERLSPEESPAQQPTPKIVGSCNCDDLTP
                     VQCHLETKELWDKFHELGTEMIITKSGRRMFPTVRVSFSGPLRQIQPADRYAVLLDVV
                     PLDSRRYRYAYHRSSWLVAGKADPPPPSRIYAHPDCPLSPEALRKQVVSFEKVKLTNN
                     EMDKSGQVVLNSMHRYQPRIHLVRLSHGQSIPGSPKELQDMDHKTFVFPETVFTAVTA
                     YQNQLITKLKIDSNPFAKGFRDSSRLSDFDRDPMDAFFFDQHMRTAPLRFFPDPLMSQ
                     LTPQEADAASMALLEKARQHLQMFGRSPYTEMLLPHLYQRSAAPPPPPPAPHLSAFQL
                     GMWQQQWPQLTAGFLASANQQAALALAAAGANRTPPPSMAVAPPAPATPTSSCGSASP
                     DLRARPQLNHYPQRFSPYQVPQHQASPPASNRAESP"
ORIGIN

Alignment Scores:
Pred. No.:                579             Length:         2563
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       2               Gaps:           0

US-09-858-332G-28 (1-6) x DMRNAH15 (1-2563)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        736 CATAATCACAATCACAAT 753

RESULT 254
AR527958
LOCUS       AR527958                2580 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 261 from patent US 6723837.
ACCESSION   AR527958
VERSION     AR527958.1  GI:53915961
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2580)
  AUTHORS   Karunanandaa,B., Yu,J. and Kishore,G.
  TITLE     Nucleic acid molecule and encoded protein associated with sterol
            synthesis and metabolism
  JOURNAL   Patent: US 6723837-A 261 20-APR-2004;
            Monsanto Technology LLC; St. Louis, MO
FEATURES             Location/Qualifiers
     source          1. .2580
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN
```

```
Alignment Scores:
Pred. No.:                583             Length:         2580
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x AR527958 (1-2580)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1900 CATAATCATAATCATAAT 1917

RESULT 255
BT022140
LOCUS       BT022140               2606 bp    mRNA    linear   INV 25-MAY-2005
DEFINITION  Drosophila melanogaster IP01538 full insert cDNA.
ACCESSION   BT022140
VERSION     BT022140.1  GI:66571139
KEYWORDS    FLI_CDNA.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 2606)
  AUTHORS   Stapleton,M., Carlson,J., Chavez,C., Frise,E., George,R.,
            Pacleb,J., Park,S., Wan,K., Yu,C. and Celniker,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAY-2005) Berkeley Drosophila Genome Project,
            Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, USA
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory
            Berkeley, CA 94720
            This clone was sequenced as part of a high-throughput process to
            sequence clones from the Drosophila Gene Collection The sequence
            has been subjected to integrity checks for sequence accuracy,
            presence of a polyA tail and contiguity within 100 kb in the
            genome. Thus we believe the sequence to reflect accurately this
            particular cDNA clone. However, there are artifacts associated with
            the generation of cDNA clones that may have not been detected in
            our initial analyses such as internal priming, priming from
            contaminating genomic DNA, retained introns due to reverse
            transcription of unspliced precursor RNAs, and reverse
            transcriptase errors that result in single base changes. For
            further information about this sequence, including its location and
            relationship to other sequences, please visit our Web site
            (http://www.fruitfly.org) or send email to cdna@fruitfly.org.
FEATURES              Location/Qualifiers
     source           1. .2606
                      /organism="Drosophila melanogaster"
                      /mol_type="mRNA"
                      /db_xref="taxon:7227"
     gene             1. .2606
                      /gene="H15"
                      /note="blastp alignment with CG6604-RA"
                      /map="25E1-25E1"
                      /db_xref="FLYBASE:FBgn0016660"
     misc_feature     1. .2605
                      /note="this clone was identified by inverse PCR using
                      templates from pooled GH, LD, LP, and SD cDNA libraries"
     misc_feature     1. .2591
                      /note="sim4 alignment with AE003609:217699-229321"
     CDS              575. .2557
                      /gene="H15"
                      /note="Longest ORF"
                      /codon_start=1
                      /product="IP01538p"
                      /protein_id="AAY51535.1"
                      /db_xref="GI:66571140"
                      /db_xref="FLYBASE:FBgn0016660"
                      /translation="MLLSNQPANTKPQQTPSPSQTQNFKSKLQQQIVSAAAAAAANIA
                      NGSSHHHHHQNHHHHHPLNNHHNHNHNQHNISFATDFSIAAIMARGGNAPSSREPSER
                      SLSPASVERYSGQDADDDVDVDVVDCSDSEMPSATAAAAAAATTAAAAAAALQAQQQA
                      RQALRVAQQQQQQQQQQRQQTHHHATTGKQQRQHHNHHSSNTNNSSNSGNSNTNSKSS
                      SQRGRSAAAVGAAATPSPPPPPPSQSPEELERLSPEESPAQQPTPKIVGSCNCDDLTP
                      VQCHLETKELWDKFHELGTEMIITKSGRRMFPTVRVSFSGPLRQIQPADRYAVLLDVV
                      PLDSRRYRYAYHRSSWLVAGKADPPPPSRIYAHPDCPLSPEALRKQVVSFEKVKLTNN
                      EMDKSGQVVLNSMHRYQPRIHLVRLSHGQSIPGSPKELQDMDHKTFVFPETVFTAVTA
                      YQNQLITKLKIDSNPFAKGFRDSSRLSDFDRDPMDAFFFDQHMRTAPLRFFPDPLMSQ
                      LTPQEADAASMALLEKARQHLQMFGRSPYTEMLLPHLYQRSAAPPPPPPAPHLSAFQL
                      GMWQQQWPQLTAGFLASANQQAALALAAAGANRTPPPSMAVAPPAPATPTSSCGSASP
                      DLRARPQLNHYPQRFSPYQVPQHQASPPASNRAESP"
ORIGIN

Alignment Scores:
Pred. No.:                588             Length:         2606
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       2               Gaps:           0

US-09-858-332G-28 (1-6) x BT022140 (1-2606)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        770 CATAATCACAATCACAAT 787

RESULT 256
AR439824
LOCUS       AR439824               2663 bp    DNA     linear   PAT 20-FEB-2004
DEFINITION  Sequence 47 from patent US 6664446.
ACCESSION   AR439824
VERSION     AR439824.1  GI:42665789
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2663)
  AUTHORS   Heard,J., Broun,P., Riechmann,J.L., Keddie,J., Pineda,O., Adam,L.,
            Samaha,R., Zhang,J., Yu,G.-L., Ratcliffe,O., Pilgrim,M.,
            Jiang,C.-Z. and Reuber,L.
  TITLE     Transgenic plants comprising polynucleotides encoding transcription
            factors that confer disease tolerance
  JOURNAL   Patent: US 6664446-A 47 16-DEC-2003;
            Mendel Biotechnology, Inc.; Hayward, CA
FEATURES              Location/Qualifiers
     source           1. .2663
                      /organism="unknown"
                      /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:                601             Length:         2663
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x AR439824 (1-2663)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        800 CACAACCACAACCACAAC 817

RESULT 257
AY144597
LOCUS       AY144597               2688 bp    mRNA    linear   INV 18-AUG-2003
DEFINITION  Dictyostelium discoideum clathrin-adaptor gamma chain Ap1g1 (ap1g1)
            mRNA, complete cds.
ACCESSION   AY144597
```

```
VERSION     AY144597.1  GI:27462057
KEYWORDS    .
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 2688)
  AUTHORS   Lefkir,Y., De Chassey,B., Dubois,A., Bogdanovic,A., Brady,R.J.,
            Destaing,O., Bruckert,F., O'Halloran,T.J., Cosson,P. and
            Letourneur,F.
  TITLE     The AP-1 Clathrin-adaptor Is Required for Lysosomal Enzymes Sorting
            and Biogenesis of the Contractile Vacuole Complex in Dictyostelium
            Cells
  JOURNAL   Mol. Biol. Cell 14 (5), 1835-1851 (2003)
   PUBMED   12802059
REFERENCE   2  (bases 1 to 2688)
  AUTHORS   Lefkir,Y. and Letourneur,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-2002) IBCP, CNRS/UMR5086, 7 Passage du Vercors,
            Lyon 69367, France
FEATURES             Location/Qualifiers
     source          1. .2688
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /db_xref="taxon:44689"
                     /cell_line="AX4"
     gene            1. .2688
                     /gene="ap1g1"
                     /note="apg1"
     CDS             1. .2688
                     /gene="ap1g1"
                     /note="Apg1"
                     /codon_start=1
                     /product="clathrin-adaptor gamma chain Ap1g1"
                     /protein_id="AAN41659.1"
                     /db_xref="GI:27462058"
                     /translation="MSSKLRDLIKTVRSCKTAAEERSQIAKESALIRTAMKEEDLESR
                     QRNVAKLLYIHMLGYPTQFGQMECLKLIVSPSYADKRIGYLGLMLLLDEKQEVLLLAT
                     NCIRGDIMNSNQFIVGVSLCAFGNICSTAMARDISPEIEKVISHSNPYIRKKAALCAI
                     RVLRKVPDLTENYIPKIKALLSERNHAVILTALTLIIEICEMDSTQIIHFKKMVPQLV
                     RILKSLTSSGYLPEHDIGGVTDPFLQVKILRLLRILGQNDPEASDAMNDILAQVSTNT
                     DSTKNVGNAILYECVQTIMTIESENGLKVMAINILGRFLLNRDNNIRYVALNTLSRVV
                     NTDIQAVQRHRNTIVECLKDPDVSIRCRALDLIYSLVTESNIRVLVRELLNFLLIADA
                     QFKSELVAKLCIVTEKYAPNKRWQIDTILRVMSIAGNFIPDEVPSNLIQLISSTPELS
                     SYAVQKLYLALKQDITQQPLTQVGLWCIGEYGDLLVADKSQLPKDEDGLSLNVSEQAV
                     IDIIDLIFRHATTTQATRQYSLTSLAKLSSRFSQSSLQRIKTMIDNYKQNINLELQQR
                     ACEYSTLFDFDKKASILDRMPPIEKQEESPHIGNKNIPTQTPPQQHYQQQQQQPQQQS
                     SQFGSILDGLDSPTQSSANSGNNNNNNNKQGGNAMSLLEDIFGSAPTPTSNGNMNNNN
                     NMNNMNNNMNNNYAMGGMGMNNNNNNSMGGMMNNNNNNNNNNNNNNNNNKSQASALLD
                     IMGDLQLTPTPQQPQSQSQQALSPTNQTSVLQPVPQPLTFLVYQKHGLNISYECSKPQ
                     PNNLSLTNINMVITNTGSSPITNFSLQAAVPKYLKIQLLAPSSTVIPPNNSGEVTQVS
                     KVLNSQQGQKPILLRLKLDFQINGQPFSDVPDTPLPSLF"
ORIGIN

Alignment Scores:
Pred. No.:              606             Length:         2688
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-858-332G-28 (1-6) x AY144597 (1-2688)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       2261 CACAATCACAATCACAAC 2278

RESULT 258
AX073210
LOCUS       AX073210               2690 bp    DNA     linear   PAT 25-JAN-2001
DEFINITION  Sequence 321 from Patent WO0102550.
ACCESSION   AX073210
VERSION     AX073210.1  GI:12583400
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1
  AUTHORS   Contreras,R.H., de Backer,M.D., Luyten,W.H., Malcorps,I.K.,
            Nelissen,B.J. and Reekmans,R.J.
  TITLE     Cell death related drug targets in yeast and fungi
  JOURNAL   Patent: WO 0102550-A 321 11-JAN-2001;
            JANSSEN PHARMACEUTICA N.V. (BE)
FEATURES             Location/Qualifiers
     source          1. .2690
                     /organism="Candida albicans"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:5476"
ORIGIN

Alignment Scores:
Pred. No.:              607             Length:         2690
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x AX073210 (1-2690)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1800 CACAACCACAACCACAAT 1817

RESULT 259
AR546931
LOCUS       AR546931               2727 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 2062 from patent US 6747137.
ACCESSION   AR546931
VERSION     AR546931.1  GI:53940106
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2727)
  AUTHORS   Weinstock,K.G. and Bush,D.
  TITLE     Nucleic acid sequences relating to Candida albicans for diagnostics
            and therapeutics
  JOURNAL   Patent: US 6747137-A 2062 08-JUN-2004;
            Genome Therapeutics Corporation; Waltham, MA
FEATURES             Location/Qualifiers
     source          1. .2727
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              615             Length:         2727
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x AR546931 (1-2727)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1829 CACAATCACAACCACAAC 1846

RESULT 260
MMU95169/c
LOCUS       MMU95169               2757 bp    mRNA    linear   ROD 07-MAY-1997
DEFINITION  Mus musculus Mtrp1 delta variant (Mtrp1) mRNA sequence.
ACCESSION   U95169
```

```
VERSION     U95169.1  GI:2052506
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2757)
  AUTHORS   Sakura,H. and Ashcroft,F.M.
  TITLE     Identification of four trp1 gene variants murine pancreatic
            beta-cells
  JOURNAL   Diabetologia 40 (5), 528-532 (1997)
   PUBMED   9165220
REFERENCE   2  (bases 1 to 2757)
  AUTHORS   Sakura,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-1996) Laboratory of Physiology, University of
            Oxford, Parks Road, Oxford OX1 3PT, UK
FEATURES             Location/Qualifiers
     source          1. .2757
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /cell_line="MIN6"
                     /tissue_type="insulinoma"
     misc_feature    472. .2644
                     /note="Mtrp1 delta; deletion variant of Mtrp1 alpha
                     (GenBank Accession Number U73625); this interval, which
                     would encode the Mtrp1 delta protein, is 257 nucleotides
                     shorter than Mtrp1 alpha due to an internal deletion and
                     may not encode a product due to a resulting frameshift"
ORIGIN

Alignment Scores:
Pred. No.:                621           Length:         2757
Score:                    42.00         Matches:        6
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              100.00%       Indels:         0
DB:                       9             Gaps:           0

US-09-858-332G-28 (1-6) x MMU95169 (1-2757)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       2170 CACAATCACAACCACAAC 2153

RESULT 261
DDI301670
LOCUS       DDI301670               2790 bp    mRNA    linear   INV 07-FEB-2003
DEFINITION  Dictyostelium discoideum mRNA for STATc transcriptional repressor
            (dstC gene).
ACCESSION   AJ301670
VERSION     AJ301670.1  GI:13235234
KEYWORDS    dstC gene; STATc protein; transcriptional repressor.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1
  AUTHORS   Fukuzawa,M., Araki,T., Adrian,I. and Williams,J.G.
  TITLE     Tyrosine phosphorylation-independent nuclear translocation of a
            dictyostelium STAT in response to DIF signaling
  JOURNAL   Mol. Cell 7 (4), 779-788 (2001)
   PUBMED   11336701
REFERENCE   2  (bases 1 to 2790)
  AUTHORS   Williams,J.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-NOV-2000) Williams J.G., School of Life Sciences,
            University Of Dundee, MSI/WTB Dow St, Dundee DD1 5EH, UNITED
            KINGDOM
FEATURES             Location/Qualifiers
     source          1. .2790
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /variety="NC4"
                     /cultivar="Ax2"
                     /db_xref="taxon:44689"
     gene            1. .2790
                     /gene="dstC"
     CDS             1. .2790
                     /gene="dstC"
                     /function="transcriptional repressor"
                     /codon_start=1
                     /evidence=experimental
                     /product="STATc protein"
                     /protein_id="CAC33514.1"
                     /db_xref="GI:13235235"
                     /db_xref="GOA:Q9BLX2"
                     /db_xref="UniProt/TrEMBL:Q9BLX2"
                     /translation="MSNNNPKKRPLDAISNTFEVKQEEPEFSSDGFNTTNDDLMSLMT
                     FLDNGTGQQNQQNQQNQQPQPQPQPQPQLPQPQSQQPIYNSNTTTVTVKTEGIATSPL
                     SNASSPISTNNNIYNNTNNNNNNNNNNNNNNNNNNNNNNNNNTATPPAIGVQQNSNI
                     PYSYPIYTDVTGQQTQHQQNIGQNSVNIDPYFQTIDGAQIQQQQLLQQQLQPIQQVNN
                     PQIDQAQIQQQQAQQIQQQQAQIQAQQAQIQQQQLEQQHLQQQQFQFQQQQQQQQQQQ
                     QQQQQQQQQQQQQQQQQQQQQQHHQQQQQQQQHQQQQQHQQQQQHQQHQQQQHQQHQQI
                     HQNQLNHQNQLNNQNQLNNQNQLNNQNQLNHQNQLNQPTQPQMQQIELPILHQLPVEP
                     QHLHPIPTHINNNGNNNNNNNNNGSNSSNSSNGIGSPDDIIEPNILSSIQHCNNNNLP
                     LPDHLLINTPYGNVLQPHQQIINECLKLHLAQKEQLDKMKIVQKQVLAHPQKETFQML
                     DNEQNTLKKQIDAEITSLQQIDQTFVLSPPEIRNVIFLLHELTIQSIQLELYHEELQL
                     LVRPQNPPPTIAALVVIEQPFPMVITKCKPLEDDPVVVQLLCGTRTELQMIGKVRATM
                     IVENQQGSKTSSSPKTIETEVVSMDETNRLAKYHLKFLNGTRKNPVTLKFGMQVQVVG
                     GTAVNIESPPTSPFIVITNECQYEESDGTLLKKDSFGNNNEIPWASYANKLQRHFLRA
                     TRQDFMKPTRYLSRHELMYIHHQFFGSKPMIPQSSFDSFWIWFGKGLQKLRYQRHVCS
                     MWQSGLIYGFISRQSVEEALRNEEQGTFLIRFSERHAGHFAVGYKVDDPDPEKRIRHY
                     LVKADDTAGAKKTLPDFLSECPQFTKILQLTIDVSTGEPRLRNFPKDVVLEPYYSKRE
                     TLPATNGYDSLPTIL"
ORIGIN

Alignment Scores:
Pred. No.:                628           Length:         2790
Score:                    42.00         Matches:        6
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              100.00%       Indels:         0
DB:                       2             Gaps:           0

US-09-858-332G-28 (1-6) x DDI301670 (1-2790)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        188 CACAACCACAACCACAAC 205

RESULT 262
AX488814
LOCUS       AX488814                2838 bp    DNA     linear   PAT 16-AUG-2002
DEFINITION  Sequence 6114 from Patent WO02053728.
ACCESSION   AX488814
VERSION     AX488814.1  GI:22322826
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1
  AUTHORS   Roemer,T., Jiang,B., Boone,C., Bussey,H. and Ohlsen,K.L.
  TITLE     Gene disruption methodologies for drug target discovery
  JOURNAL   Patent: WO 02053728-A 6114 11-JUL-2002;
            Elitra Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .2838
                     /organism="Candida albicans"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:5476"
ORIGIN

Alignment Scores:
```

```
Pred. No.:              639           Length:        2838
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-28 (1-6) x AX488814 (1-2838)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1838 CACAACCACAATCACAAT 1855

RESULT 263
YLI250175/c
LOCUS       YLI250175               2842 bp    DNA     linear   PLN 13-OCT-2000
DEFINITION  Yarrowia lipolytica MTPIII and MTPIV genes for metallothioneins.
ACCESSION   AJ250175
VERSION     AJ250175.1  GI:10801032
KEYWORDS    metallothionein; MTPIII gene; MTPIV gene.
SOURCE      Yarrowia lipolytica
  ORGANISM  Yarrowia lipolytica
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Dipodascaceae; Yarrowia.
REFERENCE   1
  AUTHORS   Garcia,S.
  JOURNAL   Thesis (1999) Departamento de Microbiologia y Genetica, Universidad
            de Salamanca, SPAIN
REFERENCE   2
  AUTHORS   Garcia,S., Prado,M. and Dominguez,A.
  TITLE     Divergence between mechanisms for metal resistance in yeasts:the
            metallothionein genes from Yarrowia lipolytica
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 2842)
  AUTHORS   Garcia,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-1999) Garcia S., Departamento De Microbiologia Y
            Genetica, Universidad de Salamanca, Edificio Departamental. Avda.
            Campo Charro s/n, Salamanca 37007, SPAIN
FEATURES             Location/Qualifiers
     source          1. .2842
                     /organism="Yarrowia lipolytica"
                     /mol_type="genomic DNA"
                     /strain="JM12"
                     /db_xref="taxon:4952"
                     /chromosome="III-IV"
     repeat_region   128. .141
                     /note="(GT)n"
                     /rpt_unit="128. .129"
     gene            complement(1555. .1796)
                     /gene="MTPIII"
     CDS             complement(1555. .1722)
                     /gene="MTPIII"
                     /function="metal quelating protein"
                     /citation=[1]
                     /codon_start=1
                     /product="putative metallothionein"
                     /protein_id="CAC12964.1"
                     /db_xref="GI:10801033"
                     /db_xref="GOA:Q9HFD0"
                     /db_xref="UniProt/TrEMBL:Q9HFD0"
                     /translation="MEFTTAMLGASLISTTSTQSKHNLVNNCCCSSSTSESSMPASCA
                     CTKCGCKTCKC"
     repeat_region   1736. .1767
                     /note="(GT)n(GA)m"
                     /rpt_unit="1736. .1737"
                     /rpt_unit="1756. .1757"
     TATA_signal     complement(1790. .1796)
                     /gene="MTPIII"
                     /note="conserved between MTPI and MTPIII"
     gene            2526. .2783
                     /gene="MTPIV"
     TATA_signal     2526. .2532
                     /gene="MTPIV"
                     /note="conserved between MTPII and MTPIV"
     CDS             2619. .2783
                     /gene="MTPIV"
                     /function="metal quelating protein"
                     /codon_start=1
                     /product="putative metallothionein"
                     /protein_id="CAC12965.1"
                     /db_xref="GI:10801034"
                     /db_xref="GOA:Q9HFC9"
                     /db_xref="UniProt/TrEMBL:Q9HFC9"
                     /translation="MEFLNANFGASLIQSKHKTTKKHNLVNSCCCSKPAEKPTNSCTC
                     SKCACDSCKC"
ORIGIN

Alignment Scores:
Pred. No.:              640           Length:        2842
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     15            Gaps:          0

US-09-858-332G-28 (1-6) x YLI250175 (1-2842)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1738 CACAACCACAACCACAAT 1721

RESULT 264
SCYLR054C/c
LOCUS       SCYLR054C               2845 bp    DNA     linear   PLN 11-AUG-1997
DEFINITION  S.cerevisiae chromosome XII reading frame ORF YLR054c.
ACCESSION   Z73226 Y13138
VERSION     Z73226.1  GI:1360393
KEYWORDS    .
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 2845)
  AUTHORS   Andre,B. and Urrestarazu,L.A.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2845)
  AUTHORS   MIPS.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAY-1996) Data collected by MIPS on behalf of the
            European yeast chromosome XII sequencing project. MIPS at the
            Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
            Martinsried, FRG; E-mail: Mewes@mips.embnet.org
FEATURES             Location/Qualifiers
     source          1. .2845
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4932"
                     /chromosome="XII"
     CDS             complement(291. .1829)
                     /note="unnamed protein product; ORF YLR054c"
                     /codon_start=1
                     /protein_id="CAA97584.1"
                     /db_xref="GI:1360394"
                     /db_xref="SGD:S0004044"
                     /db_xref="UniProt/TrEMBL:Q12202"
                     /translation="MKSKLILDPKQMALKMWELLIPKISLNILSIIYEQFDYEKMLEN
                     KSTEIIFKDLVKELLGICFAQCGSKIARFLLVKSQGEEEINFGKIVEYCKGKKLQLIN
                     STANEHPEYLSLPFEPYCFYHRFEYPAQILLHQPIILAKKYGVSCSNLNFLYGFYTRL
                     LTLSGLSINGGRCEHALSMLDSRIGGGINVASGINSGQDFTDGDNEDQDKGKNRVDKN
                     VKEGSFISLTQRFTMTSPLGVNDPALPADLEKLYLGAEYISNCDANTSGGQKRVKSPT
                     KADTGYDDKHLPDDTIHTFEDEYLADEDDFSCLGIDKRSSTKPTKPLEKFGVVAVPHF
                     IRRFSIKRSSKDKSNDMKRPYTTSSLELQLRSNHFMFAKEYQDLHRQLYYEVKPRTQS
                     ELDARRRNYSELESQMWKIKHRFNIHRGALPRPRTNPYELLLDHIDVLNRGNTGDILR
                     FTTSRYGGVDTYDSILRDQSTIMELLDKRCAYSPPVLFDNEAKERDHDRDCYQDYNHD
                     HRSH"
```

```
ORIGIN

Alignment Scores:
Pred. No.:                640             Length:         2845
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       15              Gaps:           0

US-09-858-332G-28 (1-6) x SCYLR054C (1-2845)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       2828 CATAACCATAATCATAAT 2811

RESULT 265
MMU95168/c
LOCUS       MMU95168                2859 bp    mRNA    linear   ROD 07-MAY-1997
DEFINITION  Mus musculus Mtrp1 gamma variant (Mtrp1) mRNA sequence.
ACCESSION   U95168
VERSION     U95168.1  GI:2052505
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2859)
  AUTHORS   Sakura,H. and Ashcroft,F.M.
  TITLE     Identification of four trp1 gene variants murine pancreatic
            beta-cells
  JOURNAL   Diabetologia 40 (5), 528-532 (1997)
   PUBMED   9165220
REFERENCE   2  (bases 1 to 2859)
  AUTHORS   Sakura,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-1996) Laboratory of Physiology, University of
            Oxford, Parks Road, Oxford OX1 3PT, UK
FEATURES             Location/Qualifiers
     source          1. .2859
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /cell_line="MIN6"
                     /tissue_type="insulinoma"
     misc_feature    472. .2746
                     /note="Mtrp1 gamma; deletion variant of Mtrp1 alpha
                     (GenBank Accession Number U73625); this interval, which
                     would encode the Mtrp1 gamma protein, is 155 nucleotides
                     shorter than Mtrp1 alpha due to an internal deletion and
                     may not encode a product due to a resulting frameshift"
ORIGIN

Alignment Scores:
Pred. No.:                643             Length:         2859
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       9               Gaps:           0

US-09-858-332G-28 (1-6) x MMU95168 (1-2859)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       2272 CACAATCACAACCACAAC 2255

RESULT 266
AY051856
LOCUS       AY051856                2874 bp    mRNA    linear   INV 27-AUG-2001
DEFINITION  Drosophila melanogaster LD35413 full length cDNA.
ACCESSION   AY051856
VERSION     AY051856.1  GI:15292022
KEYWORDS    FLI_CDNA.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 2874)
  AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J.,
            Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George,R.,
            Gonzalez,M., Guarin,H., Li,P., Liao,G., Miranda,A., Mungall,C.J.,
            Nunoo,J., Pacleb,J., Paragas,V., Park,S., Phouanenavong,S., Wan,K.,
            Yu,C., Lewis,S.E., Rubin,G.M. and Celniker,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-AUG-2001) Berkeley Drosophila Genome Project,
            Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, USA
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory
            Berkeley, CA 94720
            This clone was sequenced as part of a high-throughput process to
            sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
            Science 2000). The sequence has been subjected to integrity checks
            for sequence accuracy, presence of a polyA tail and contiguity
            within 100 kb in the genome. Thus we believe the sequence to
            reflect accurately this particular cDNA clone.  However, there are
            artifacts associated with the generation of cDNA clones that may
            have not been detected in our initial analyses such as internal
            priming, priming from contaminating genomic DNA, retained introns
            due to reverse transcription of unspliced precursor RNAs, and
            reverse transcriptase errors that result in single base changes.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our Web site
            (http://fruitfly.berkeley.edu) or send email to
            cdna@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1. .2874
                     /organism="Drosophila melanogaster"
                     /mol_type="mRNA"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /map="87A6-87A6"
     gene            1. .2874
                     /gene="CG14732"
                     /note="alignment with genomic scaffold AE003693. gene does
                     not completely overlap longest ORF"
                     /db_xref="FLYBASE:FBgn0037965"
     CDS             109. .2733
                     /gene="CG14732"
                     /note="Longest ORF"
                     /codon_start=1
                     /product="LD35413p"
                     /protein_id="AAK93280.1"
                     /db_xref="GI:15292023"
                     /db_xref="FLYBASE:FBgn0037965"
                     /translation="MFPGGNGNADLGYNKHAHVHGLGQHGHGHNHNHNQFLAQPPPPP
                     THFSLPSGGGQGMVTPMVAAGLPLAMQGGVGIDWAQLAQQWIHMRDATPVPMPLAPPP
                     PIISNLREYHQQTLAMPVPSVVRAGFPQLEEHGEADMDMDDENDRGHSTETPPPPAPL
                     VTQSQWLAGTEVAGQIPDASNGAAVTLGKQPWTGWQMQTDKSGNSTAHIPSLLKLNVS
                     NPNEPHQQLQLQLQAHQQQQQQHAPQPTHPHHPPAHQPHAHPHHLTPHTHLPPQHTQQ
                     THAESGSSGASSEIDANKRKMLPAWIREGLEKMEREKQRQLERQQSSMTTPDVEVNHV
                     KQVSSKTLTTPGNLLNIANVASDSEDSIDVPVEARGELKVQQIGNELISKDDDRSSSS
                     SEEPESELHGAIETGNRTVAEHVRMATVNDVNGKNYEERLADLMLVVRRTLTEILLET
                     TNEEIAAIAGETLKAHRAKASSAQVIRKSALSSITGNLGLAAYGDSSSETEDDEDDED
                     ERQAGAGKDAEKSAQLSAEELKARIRRSKRSFEKVIDDIEDRVAKQELLDEQTLLRHR
                     KRELERSVTGGGEHRRPAANPEPPSESASQATQEKHQQSNGKRLSRKERTTRFSDNKD
                     GKQQSQSFVQQVVATAVVPPPGSLSNSSPQKLKPVPNPATNLLQMPESVATMLTAADK
                     ALHKANKSSHKKSKRRHSSPSSSSSSSGSSSSSDSDDSSSTSSGSKTSGSRSKRSSRH
                     GHRSHHSSSRSKYERRDRDRERERARERERDRDRSHRQHRSSQLSGSHHQKQRRHRES
                     SHSGEDAGGSSSYHRSSRQSSSRSHDHGSNRSHDHGSSSSGRKRHRTRSRSKSRSTHH
                     SSSKAYSSTSASHPRKRH"
ORIGIN
```

```
Alignment Scores:
Pred. No.:              647           Length:        2874
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     2             Gaps:          0

US-09-858-332G-28 (1-6) x AY051856 (1-2874)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        193 CACAACCATAATCACAAT 210

RESULT 267
MMU95167/c
LOCUS       MMU95167                2912 bp    mRNA    linear   ROD 07-MAY-1997
DEFINITION  Mus musculus Mtrp1 beta variant (Mtrp1) mRNA, complete cds.
ACCESSION   U95167
VERSION     U95167.1  GI:2052503
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2912)
  AUTHORS   Sakura,H. and Ashcroft,F.M.
  TITLE     Identification of four trp1 gene variants murine pancreatic
            beta-cells
  JOURNAL   Diabetologia 40 (5), 528-532 (1997)
   PUBMED   9165220
REFERENCE   2  (bases 1 to 2912)
  AUTHORS   Sakura,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-1996) Laboratory of Physiology, University of
            Oxford, Parks Road, Oxford OX1 3PT, UK
FEATURES             Location/Qualifiers
     source          1. .2912
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /cell_line="MIN6"
                     /tissue_type="insulinoma"
     gene            1. .2912
                     /gene="Mtrp1"
     CDS             472. .2799
                     /gene="Mtrp1"
                     /function="channel for calcium release activated current"
                     /note="deletion variant of Mtrp1 alpha (GenBank Accession
                     Number U73625); the beta variant is 34 amino acids shorter
                     than the alpha form due to an internal deletion"
                     /codon_start=1
                     /product="Mtrp1 beta"
                     /protein_id="AAC53162.1"
                     /db_xref="GI:2052504"
                     /translation="MGAPPPSPGLPPSWAAMMAALYPSTDLSGVSSSSLPSSPSSSSP
                     NEVMALKDVREVKEENTLNEKLFLLACDKGDYYMVKKILEENSSGDLNINCVDVLGRN
                     AVTITIENESLDILQLLLDYGCQKLMERIQNPEYSTTMDVAPVILAAHRNNYEILTML
                     LKQDVSLPKPHAVGCECTLCSAKNKKDSLRHSRFRLDIYRCLASPALIMLTEEDPILR
                     AFELSADLKELSLVEVEFRNDYEELARQCKMFAKDLLAQARNSRELEVILNHTSSDEP
                     LDKRGLLEERMNLSRLKLAIKYNQKEFVSQSNCQQFLNTVWFGQMSGYRRKPTCKKIM
                     TVLTVGIFWPVLSLCYLIAPKSQFGRIIHTPFMKFIIHGASYFTFLLLLNLYSLVYNE
                     DKKNTMGPALERIDYLLILWIIGMIWSDIKRLWYEGLEDFLEESRNQLSFVMNSLYLA
                     TFALKVVAHNKFHDFADRKDWDAFHPTLVAEGLFAFANVLSYLRLFFMYTTSSILGPL
                     QISMGQMLQDFGKFLGMFLLVLFSFTIGLTQLYDKGYTSKEQKDCVGIFCEQQSNDTF
                     HSFIGTCFALFWYIFSLAHVAIFVTRFSYGEELQSFVGAVIVGTYNVVVVIVLTKLLV
                     AMLHKSFQLIANHEDKEWKFARAKLWLSYFDDKCTLPPPFNIIPSPKTICYMISSLSK
                     WICSHTSKGKVKRQNSLKEWRNLKQKRDENYQKVMCCLVHRYLTSMRQKMQSTDQATV
                     ENLNELRQDLSKFRNEIRDLLGFRTSKYAMFYPRN"
ORIGIN
```

```
Alignment Scores:
Pred. No.:              655           Length:        2912
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     9             Gaps:          0

US-09-858-332G-28 (1-6) x MMU95167 (1-2912)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       2325 CACAATCACAACCACAAC 2308

RESULT 268
AX713676/c
LOCUS       AX713676                2959 bp    DNA     linear   PAT 15-APR-2003
DEFINITION  Sequence 360 from Patent EP1293569.
ACCESSION   AX713676
VERSION     AX713676.1  GI:29888534
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,
            Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R.,
            Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahari,K. and
            Masuho,Y.
  TITLE     Full-length cDNAs
  JOURNAL   Patent: EP 1293569-A 360 19-MAR-2003;
            Helix Research Institute (JP) ; Research Association for
            Biotechnology (JP)
FEATURES             Location/Qualifiers
     source          1. .2959
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:              665           Length:        2959
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-28 (1-6) x AX713676 (1-2959)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       2769 CATAATCATAATCATAAT 2752

RESULT 269
AK055362/c
LOCUS       AK055362                2959 bp    mRNA    linear   PRI 30-JAN-2004
DEFINITION  Homo sapiens cDNA FLJ30800 fis, clone FEBRA2001197.
ACCESSION   AK055362
VERSION     AK055362.1  GI:16550074
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Ota,T., Suzuki,Y., Nishikawa,T., Otsuki,T., Sugiyama,T., Irie,R.,
            Wakamatsu,A., Hayashi,K., Sato,H., Nagai,K., Kimura,K., Makita,H.,
            Sekine,M., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
            Ishii,S., Yamamoto,J., Saito,K., Kawai,Y., Isono,Y., Nakamura,Y.,
```

Nagahari,K., Murakami,K., Yasuda,T., Iwayanagi,T., Wagatsuma,M., Shiratori,A., Sudo,H., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H., Sugawara,M., Takahashi,M., Kanda,K., Yokoi,T., Furuya,T., Kikkawa,E., Omura,Y., Abe,K., Kamihara,K., Katsuta,N., Sato,K., Tanikawa,M., Yamazaki,M., Ninomiya,K., Ishibashi,T., Yamashita,H., Murakawa,K., Fujimori,K., Tanai,H., Kimata,M., Watanabe,M., Hiraoka,S., Chiba,Y., Ishida,S., Ono,Y., Takiguchi,S., Watanabe,S., Yosida,M., Hotuta,T., Kusano,J., Kanehori,K., Takahashi-Fujii,A., Hara,H., Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,R., Takeuchi,K., Arita,M., Imose,N., Musashino,K., Yuuki,H., Oshima,A., Sasaki,N., Aotsuka,S., Yoshikawa,Y., Matsunawa,H., Ichihara,T., Shiohata,N., Sano,S., Moriya,S., Momiyama,H., Satoh,N., Takami,S., Terashima,Y., Suzuki,O., Nakagawa,S., Senoh,A., Mizoguchi,H., Goto,Y., Shimizu,F., Wakebe,H., Hishigaki,H., Watanabe,T., Sugiyama,A., Takemoto,M., Kawakami,B., Yamazaki,M., Watanabe,K., Kumagai,A., Itakura,S., Fukuzumi,Y., Fujimori,Y., Komiyama,M., Tashiro,H., Tanigami,A., Fujiwara,T., Ono,T., Yamada,K., Fujii,Y., Ozaki,K., Hirao,M., Ohmori,Y., Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T., Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Sasaki,M., Togashi,T., Oyama,M., Hata,H., Watanabe,M., Komatsu,T., Mizushima-Sugano,J., Satoh,T., Shirai,Y., Takahashi,Y., Nakagawa,K., Okumura,K., Nagase,T., Nomura,N., Kikuchi,H., Masuho,Y., Yamashita,R., Nakai,K., Yada,T., Nakamura,Y., Ohara,O., Isogai,T. and Sugano,S.

TITLE Complete sequencing and characterization of 21,243 full-length human cDNAs

JOURNAL Nat. Genet. 36 (1), 40-45 (2004)

PUBMED 14702039

REFERENCE 2

AUTHORS Ninomiya,K., Wagatsuma,M., Kanda,K., Kondo,H., Yokoi,T., Kodaira,H., Furuya,T., Takahashi,M., Kikkawa,E., Omura,Y., Abe,K., Kamihara,K., Katsuta,N., Sato,K., Tanikawa,M., Yamazaki,M., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Murakawa,K., Kanehori,K., Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.

TITLE NEDO human cDNA sequencing project

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 2959)

AUTHORS Isogai,T., Otsuki,T. and Sugiyama,T.

TITLE Direct Submission

JOURNAL Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)

COMMENT NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.

FEATURES Location/Qualifiers

source 1. .2959
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="FEBRA2001197"
/tissue_type="brain"
/clone_lib="FEBRA2"
/dev_stage="fetus"
/note="cloning vector: pME18SFL3"

ORIGIN

Alignment Scores:

Pred. No.:	665	Length:	2959
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	8	Gaps:	0

US-09-858-332G-28 (1-6) x AK055362 (1-2959)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       2769 CATAATCATAATCATAAT 2752
```

RESULT 270
MMU73625/c
LOCUS MMU73625 3014 bp mRNA linear ROD 28-MAR-1997
DEFINITION Mus musculus Mtrp1 alpha (Mtrp1) mRNA, complete cds.
ACCESSION U73625
VERSION U73625.1 GI:1825486
KEYWORDS .
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 3014)
AUTHORS Sakura,H. and Ashcroft,F.M.
TITLE Identification of four trp1 gene variants in murine pancreatic beta-cells
JOURNAL Diabetologia 40 (1997) In press

REFERENCE 2 (bases 1 to 3014)
AUTHORS Sakura,H.
TITLE Direct Submission
JOURNAL Submitted (08-OCT-1996) Laboratory of Physiology, University of Oxford, Parks Road, Oxford OX1 3PT, UK

FEATURES Location/Qualifiers

source 1. .3014
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/cell_line="MIN6"
/cell_type="insulinoma"

gene 1. .3014
/gene="Mtrp1"

CDS 472. .2901
/gene="Mtrp1"
/function="channel for calcium release activated current"
/codon_start=1
/product="Mtrp1 alpha"
/protein_id="AAB50622.1"
/db_xref="GI:1911245"
/translation="MGAPPPSPGLPPSWAAMMAALYPSTDLSGVSSSSLPSSPSSSSP
NEVMALKDVREVKEENTLNEKLFLLACDKGDYYMVKKILEENSSGDLNINCVDVLGRN
AVTITIENESLDILQLLLDYGCQSADALLVAIDSEVVGAVDILLNHRPKRSSRPTIVK
LMERIQNPEYSTTMDVAPVILAAHRNNYEILTMLLKQDVSLPKPHAVGCECTLCSAKN
KKDSLRHSRFRLDIYRCLASPALIMLTEEDPILRAFELSADLKELSLVEVEFRNDYEE
LARQCKMFAKDLLAQARNSRELEVILNHTSSDEPLDKRGLLEERMNLSRLKLAIKYNQ
KEFVSQSNCQQFLNTVWFGQMSGYRRKPTCKKIMTVLTVGIFWPVLSLCYLIAPKSQF
GRIIHTPFMKFIIHGASYFTFLLLLNLYSLVYNEDKKNTMGPALERIDYLLILWIIGM
IWSDIKRLWYEGLEDFLEESRNQLSFVMNSLYLATFALKVVAHNKFHDFADRKDWDAF
HPTLVAEGLFAFANVLSYLRLFFMYTTSSILGPLQISMGQMLQDFGKFLGMFLLVLFS
FTIGLTQLYDKGYTSKEQKDCVGIFCEQQSNDTFHSFIGTCFALFWYIFSLAHVAIFV
TRFSYGEELQSFVGAVIVGTYNVVVVIVLTKLLVAMLHKSFQLIANHEDKEWKFARAK
LWLSYFDDKCTLPPPFNIIPSPKTICYMISSLSKWICSHTSKGKVKRQNSLKEWRNLK
QKRDENYQKVMCCLVHRYLTSMRQKMQSTDQATVENLNELRQDLSKFRNEIRDLLGFR
TSKYAMFYPRN"

ORIGIN

Alignment Scores:

Pred. No.:	677	Length:	3014
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

US-09-858-332G-28 (1-6) x MMU73625 (1-3014)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       2427 CACAATCACAACCACAAC 2410

RESULT 271
AX536944
LOCUS       AX536944               3042 bp    DNA     linear   PAT 22-NOV-2002
DEFINITION  Sequence 545 from Patent WO02064766.
ACCESSION   AX536944
VERSION     AX536944.1  GI:25263379
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1
  AUTHORS   Contreras,R.H., Eberhardt,I., Luyten,W.H. and Reekmans,R.J.
  TITLE     Bax-responsive genes for drug target identification in yeast and
            fungi
  JOURNAL   Patent: WO 02064766-A 545 22-AUG-2002;
            JANSSEN PHARMACEUTICA N.V. (BE)
FEATURES             Location/Qualifiers
     source          1. .3042
                     /organism="Candida albicans"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:5476"
ORIGIN

Alignment Scores:
Pred. No.:              683           Length:        3042
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-28 (1-6) x AX536944 (1-3042)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       2696 CACAATCACAATCACAAT 2713

RESULT 272
ACH251521
LOCUS       ACH251521              3089 bp    DNA     linear   PLN 31-AUG-2001
DEFINITION  Acremonium chrysogenum pacC gene for putative transcription factor,
            exons 1-4.
ACCESSION   AJ251521
VERSION     AJ251521.1  GI:14132802
KEYWORDS    pacC gene; transcription factor.
SOURCE      Acremonium chrysogenum
  ORGANISM  Acremonium chrysogenum
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Hypocreomycetidae; Hypocreales; mitosporic Hypocreales; Acremonium.
REFERENCE   1
  AUTHORS   Schmitt,E.K., Kempken,R. and Kuck,U.
  TITLE     Functional analysis of promoter sequences of cephalosporin C
            biosynthesis genes from Acremonium chrysogenum: specific
            DNA-protein interactions and characterization of the transcription
            factor PACC
  JOURNAL   Mol. Genet. Genomics 265 (3), 508-518 (2001)
   PUBMED   11405634
REFERENCE   2  (bases 1 to 3089)
  AUTHORS   Schmitt,E.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-DEC-1999) Schmitt E.K., Lehrstuhl fuer Allgemeine
            Botanik, Ruhr-Universitaet Bochum, Universtitaetsstrasse 150, 44780
            Bochum, GERMANY
FEATURES             Location/Qualifiers
     source          1. .3089
                     /organism="Acremonium chrysogenum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5044"
     gene            join(175. .238,283. .388,392. .397,574. .578,794. .1006,
                     1007. .1078,1079. .1282,1283. .1358,1359. .1446,1447. .1504,
                     1505. .2865,3089)
                     /gene="pacC"
     CAAT_signal     175. .179
                     /gene="pacC"
     CAAT_signal     206. .210
                     /gene="pacC"
     enhancer        223. .228
                     /gene="pacC"
                     /note="putative PACC binding site"
     enhancer        233. .238
                     /gene="pacC"
                     /note="putative PACC binding site"
     gene            complement(251. .256)
                     /gene="pacC"
     enhancer        complement(251. .256)
                     /gene="pacC"
                     /note="putative PACC binding site"
     enhancer        283. .388
                     /gene="pacC"
                     /note="putative PaCC binding site"
     enhancer        392. .397
                     /gene="pacC"
                     /note="putative PACC binding site"
     TATA_signal     574. .578
                     /gene="pacC"
     mRNA            join(794. .1006,1079. .1282,1359. .1446,1505. .2865)
                     /gene="pacC"
     CDS             join(794. .1006,1079. .1282,1359. .1446,1505. .2865)
                     /gene="pacC"
                     /function="pH regulation"
                     /codon_start=1
                     /product="putative transcription factor"
                     /protein_id="CAC38840.1"
                     /db_xref="GI:14132803"
                     /db_xref="UniProt/TrEMBL:Q96X49"
                     /translation="MSPPAPDAQPSSGSSSDNSNADSKSTTPPPPSSTTSTTSAAAIS
                     AAVVPDDNLTCRWNACNQKFPNPETLYDHICERHVGRKSTNNLNLTCQWNSCRTTTVK
                     RDHITSHIRVHVPLKPHKCEFCGKSFKRPQDLKKHVKTHADDSVLVRPSQDNHNQGGM
                     NYRPHLNKTPSYYDHNGQMRTQYQHHQPAHPNAYYAPQPSTNYGLYFNQPPLNNHQRT
                     EHLGYGAPPGGYDRKRAFDMVDDFFSSAKRREIDPSSYSQIGRSLLPLHGTLSIPSGP
                     MPATEQHYVPQPQHAGIAHAGPAPTQNPLAQQYYLPMPPNARTQKDLVQIDNLLGQMQ
                     DTIYENANHATAGVQIHHGHDHYSGYRSSQSPPIAQRGSPGGMPVGPDGYHQPVSAAT
                     MASPLTAMSSTGTPAVTPPSSAVSYTSGHSPSPSAPAMSPQSRHSSTSGSVMYPSLPT
                     SLPAVSQGFGQSTTTTLGPSFESNERRRYSGGMLQRARAAPPRSTDNHNHNHNHGAST
                     PKAPVSPALASPSSEVSDVSEATRREREQYDRWLENMRVIESLREYVRGRLERREYVE
                     EQENEHQGQGEAARKERDGNPMDVDARSPRSPAREQPGRPREGSSLYPILRMPGD"
     exon            794. .1006
                     /gene="pacC"
                     /number=1
     intron          1007. .1078
                     /gene="pacC"
                     /number=1
     exon            1079. .1282
                     /gene="pacC"
                     /number=2
     intron          1283. .1358
                     /gene="pacC"
                     /number=2
     exon            1359. .1446
                     /gene="pacC"
                     /number=3
     intron          1447. .1504
                     /gene="pacC"
                     /number=3
     exon            1505. .2865
                     /gene="pacC"
                     /number=4
     polyA_site      3089
                     /gene="pacC"
                     /evidence=experimental
ORIGIN
```

```
Alignment Scores:
Pred. No.:                693             Length:         3089
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       15              Gaps:           0

US-09-858-332G-28 (1-6) x ACH251521 (1-3089)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       2491 CACAACCACAACCACAAC 2508

RESULT 273
AX058035
LOCUS       AX058035               3159 bp    DNA     linear   PAT 17-JAN-2001
DEFINITION  Sequence 1 from Patent WO0077215.
ACCESSION   AX058035
VERSION     AX058035.1  GI:12310627
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1
  AUTHORS   Buurman,E.T., Davidov,E.J., Jiang,W. and Kenniston,J.A.
  TITLE     Cdc68p chromatin remodeling factor
  JOURNAL   Patent: WO 0077215-A 1 21-DEC-2000;
            Anadys Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .3159
                     /organism="Candida albicans"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:5476"
ORIGIN

Alignment Scores:
Pred. No.:                708             Length:         3159
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x AX058035 (1-3159)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       2162 CACAACCACAATCACAAT 2179

RESULT 274
YSCSPT8TF
LOCUS       YSCSPT8TF              3223 bp    DNA     linear   PLN 25-NOV-1994
DEFINITION  Saccharomyces cerevisiae transcription factor (SPT8) gene, complete
            cds.
ACCESSION   M94955
VERSION     M94955.1  GI:452154
KEYWORDS    acidic protein; transcription factor.
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 3223)
  AUTHORS   Eisenmann,D.M., Chapon,C., Roberts,S.M., Dollard,C. and Winston,F.
  TITLE     The Saccharomyces cerevisiae SPT8 gene encodes a very acidic
            protein that is functionally related to SPT3 and TATA-binding
            protein
  JOURNAL   Genetics 137 (3), 647-657 (1994)
   PUBMED   8088510
REFERENCE   2  (bases 1 to 3223)
  AUTHORS   Winston,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-FEB-1994) Fred Winston, Department of Genetics,
            Harvard Medical School, 200 Longwood Ave., Boston, MA 02115, USA
COMMENT     Original source text: Saccharomyces cerevisiae DNA.
FEATURES             Location/Qualifiers
     source          1. .3223
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4932"
     gene            721. .2529
                     /gene="SPT8"
     CDS             721. .2529
                     /gene="SPT8"
                     /note="Spt8p"
                     /codon_start=1
                     /product="transcription factor"
                     /protein_id="AAA53585.1"
                     /db_xref="GI:452155"
                     /translation="MDEVDDILINNQVVDDEEDDEEMLSGLENDSKQDLEGNDDGGED
                     EEDDDDDDEDDDDDEDERE DDDEQEDDDGEDDAARMDKTATPTNEHQHDEQKAAAAGA
                     GGAGDSGDAVTKIGSEDVKLSDVDGGVGSREASSSTHEASANGEVYEYYKHMLNAAQI
                     ADSYNIYPTAAIPIQTHVNALAVSRGLKYLFLGGSDGYIRKYDLLNTLEGKLSLTILQ
                     KHSLAESIQNAGILQSYWENEIPQKKSEMKLSANKTDYEPKVSPVHSLEVQSECLFIL
                     SGLQNGGITMQGVRYMEGSIAHYFKGRNGHTQIVNILRLNGQEDRFLSGSWDKRLLEW
                     DLQTGDIVNEFKKSRSELSSLEMRPLYSSVDVSGNVNSGKENENADDDMDSLFGDEDE
                     DEKQDAGNEPVETGDGSNGEENKEQISEESLNIVYDESVFMTSGLNGSVHIWDRRMTQ
                     SPALSLERGAGVPPWCLSACWGVDGDHVYAGRRNACVEQFDLKMPSKPIHNLKLPSIS
                     GPVSCVKAMPNNKHLLCASRDNIRLYNVEIAVDASNSTTKSSKVPFLIVPGHHGGIIS
                     NLYLDPTSRFIISTSGNRGWQGNSTDTTLIYDIDLE"
     repeat_region   1634. .1737
                     /note="transducin repeat"
ORIGIN

Alignment Scores:
Pred. No.:                722             Length:         3223
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       15              Gaps:           0

US-09-858-332G-28 (1-6) x YSCSPT8TF (1-3223)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       2547 CATAACCATAATCATAAT 2564

RESULT 275
CQ614937
LOCUS       CQ614937               3228 bp    DNA     linear   PAT 02-FEB-2004
DEFINITION  Sequence 42695 from Patent WO0171042.
ACCESSION   CQ614937
VERSION     CQ614937.1  GI:41665271
KEYWORDS    .
SOURCE      Drosophila sp.
  ORGANISM  Drosophila sp.
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1
  AUTHORS   Venter,J.C., Adams,M., Li,P.W. and Myers,E.W.
  TITLE     Detection kits, such as nucleic acid arrays, for detecting the
            expression of 10,000 or more Drosophila genes and uses thereof
  JOURNAL   Patent: WO 0171042-A 42695 27-SEP-2001;
            PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1. .3228
                     /organism="Drosophila sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:7242"
ORIGIN

Alignment Scores:
```

```
Pred. No.:              723             Length:         3228
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x CQ614937 (1-3228)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        500 CACAATCACAACCACAAT 517

RESULT 276
DMU89385
LOCUS       DMU89385               3237 bp    mRNA    linear   INV 25-FEB-2000
DEFINITION  Drosophila melanogaster LIM homeobox protein (islet) mRNA, complete
            cds.
ACCESSION   U89385
VERSION     U89385.1  GI:1895061
KEYWORDS    .
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 3237)
  AUTHORS   Thor,S. and Thomas,J.B.
  TITLE     The Drosophila islet gene governs axon pathfinding and
            neurotransmitter identity
  JOURNAL   Neuron 18 (3), 397-409 (1997)
   PUBMED   9115734
REFERENCE   2  (bases 1 to 3237)
  AUTHORS   Thor,S. and Thomas,J.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-FEB-1997) Molecular Neurobiology Lab, The Salk
            Institute, 10010 N Torrey Pines Rd., La Jolla, CA 92037, USA
FEATURES             Location/Qualifiers
     source          1. .3237
                     /organism="Drosophila melanogaster"
                     /mol_type="mRNA"
                     /db_xref="taxon:7227"
                     /chromosome="2"
                     /map="37AA"
     gene            1. .3237
                     /gene="islet"
     CDS             462. .2066
                     /gene="islet"
                     /function="transcription factor"
                     /note="contains LIM homeodomain; similar to vertebrate
                     islet-1 and islet-2 proteins"
                     /codon_start=1
                     /product="LIM homeobox protein"
                     /protein_id="AAB49892.1"
                     /db_xref="GI:1895062"
                     /translation="MVMAEIGGHLAHQLPLHNHNHNQTGLQPSLVMNHHLDLDCHGHD
                     VIKKQRLSHCVGCGGQIHDQYILRVAPDLEWHAACLKCQECRQFLDESCTCFVRDGKT
                     YCKRDYVRLFGTKCDKCGNSFSKNDFVMRAKTKIFHIECFRCSACARQLLPGDEFALR
                     DAGALYCKEDHDVLEKSSQSSLTSSSVESNNNISSSNNNNTNLSNNNHSSELGSMSDS
                     GSESGSHKSIRDKRPSGPSDGKPTRVRTVLNEKQLHTLRTCYNANPRPDALMKEQLVE
                     MTSLSPRVIRVWFQNKRCKDKKKTIQMKLQMHEEKEGRKLGYGAMQGIPMIASSPVRH
                     DSPLNLQGLDVQTYQPPWKALSDFALHADLDSNGAINTHTPAFQQLVNQMHGYDLNGM
                     PILPPHPHSHPAQGPPHQHPPPPGGPHNHQNQQPNQQPGGSSLDSGITSHHHPDSTDS
                     YVTYLESDDKSKLALTPSSSSSASAGTSISSPPSGVGAGGGGAVGGGSGVLGLGVVAN
                     QSATEQLMQMLQKVTGSASPASHAVL"
ORIGIN

Alignment Scores:
Pred. No.:              725             Length:         3237
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-858-332G-28 (1-6) x DMU89385 (1-3237)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        510 CACAATCACAACCACAAT 527

RESULT 277
CQ850350/c
LOCUS       CQ850350               3277 bp    DNA     linear   PAT 23-AUG-2004
DEFINITION  Sequence 819 from Patent EP1447413.
ACCESSION   CQ850350
VERSION     CQ850350.1  GI:51508562
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Isogai,T., Yamamoto,J., Nishikawa,T., Isono,Y., Sugiyama,T.,
            Otsuki,T., Wakamatsu,A., Ishii,S., Nagai,K. and Irie,R.
  TITLE     Full-length human cDNA
  JOURNAL   Patent: EP 1447413-A 819 18-AUG-2004;
            Research Association for Biotechnology (JP)
FEATURES             Location/Qualifiers
     source          1. .3277
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:              734             Length:         3277
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x CQ850350 (1-3277)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       2773 CATAATCATAATCATAAT 2756

RESULT 278
AK127474/c
LOCUS       AK127474               3277 bp    mRNA    linear   PRI 19-FEB-2004
DEFINITION  Homo sapiens cDNA FLJ45566 fis, clone BRTHA3009858.
ACCESSION   AK127474
VERSION     AK127474.1  GI:34534402
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Kanehori,K., Ishibashi,T., Chiba,Y., Fujimori,K., Hiraoka,S.,
            Tanai,H., Watanabe,S., Ishida,S., Ono,Y., Hotuta,T., Watanabe,M.,
            Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
            Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
            Kimura,K., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
            Wagatsuma,M., Takahashi-Fujii,A., Oshima,A., Sugiyama,A.,
            Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K.
            and Isogai,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3277)
  AUTHORS   Isogai,T. and Yamamoto,J.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (15-JUL-2003) Takao Isogai, FLJ Project(HRI Team); 2-6-7
            Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
  COMMENT   NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: HRI and
            RAB; annotation: Reverse Proteomics Research Institute, HRI and
            RAB.
  FEATURES             Location/Qualifiers
       source          1. .3277
                       /organism="Homo sapiens"
                       /mol_type="mRNA"
                       /db_xref="taxon:9606"
                       /clone="BRTHA3009858"
                       /tissue_type="thalamus"
                       /clone_lib="BRTHA3"
                       /note="cloning vector: pME18SFL3"
  ORIGIN

  Alignment Scores:
  Pred. No.:                734             Length:         3277
  Score:                    42.00           Matches:        6
  Percent Similarity:       100.00%         Conservative:   0
  Best Local Similarity:    100.00%         Mismatches:     0
  Query Match:              100.00%         Indels:         0
  DB:                       8               Gaps:           0

  US-09-858-332G-28 (1-6) x AK127474 (1-3277)

  Qy          1 HisAsnHisAsnHisAsn 6
                ||||||||||||||||||
  Db       2773 CATAATCATAATCATAAT 2756

  RESULT 279
  AF145674
  LOCUS       AF145674               3317 bp    mRNA    linear   INV 14-JUN-1999
  DEFINITION  Drosophila melanogaster clone GH12341 tailup (tup) mRNA, complete
              cds.
  ACCESSION   AF145674
  VERSION     AF145674.1  GI:5052637
  KEYWORDS    FLI_CDNA.
  SOURCE      Drosophila melanogaster (fruit fly)
    ORGANISM  Drosophila melanogaster
              Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
              Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
              Ephydroidea; Drosophilidae; Drosophila.
  REFERENCE   1  (bases 1 to 3317)
    AUTHORS   Rubin,G.M., Wan,K.H., Harvey,D., Lewis,S.E., Brokstein,P.,
              Tsang,G., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
              Butenhoff,C., Champe,M., Chavez,C., Chew,M., Doyle,C.M.,
              Farfan,D.E., Frise,E., Galle,R., George,R.A., Harris,N.L.,
              Hoskins,R.A., Evans-Holm,M., Houston,K.A., Hummasti,S.R., Kim,E.,
              Li,P., Moshrefi,M., Pacleb,J.M., Park,S., Sequeira,A., Sethi,H.,
              Snir,E., Svirskas,R.R., Weinburg,T. and Celniker,S.E.
    TITLE     Full length Drosophila melanogaster cDNA sequence
    JOURNAL   Unpublished
  REFERENCE   2  (bases 1 to 3317)
    AUTHORS   Rubin,G.M., Wan,K.H., Harvey,D., Lewis,S.E., Brokstein,P.,
              Tsang,G., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
              Butenhoff,C., Champe,M., Chavez,C., Chew,M., Doyle,C.M.,
              Farfan,D.E., Frise,E., Galle,R., George,R.A., Harris,N.L.,
              Hoskins,R.A., Evans-Holm,M., Houston,K.A., Hummasti,S.R., Kim,E.,
              Li,P., Moshrefi,M., Pacleb,J.M., Park,S., Sequeira,A., Sethi,H.,
              Snir,E., Svirskas,R.R., Weinburg,T. and Celniker,S.E.
    TITLE     Direct Submission
    JOURNAL   Submitted (23-APR-1999) Berkeley Drosophila Genome Project,
              University of California Berkeley, Berkeley, CA 94720, USA
  COMMENT     For further information about this sequence, including its location
              and relationship to other sequences, please visit our web site
              (http://fruitfly.berkeley.edu) or send email to
              cdna@fruitfly.berkeley.edu.
  FEATURES             Location/Qualifiers
       source          1. .3317
                       /organism="Drosophila melanogaster"
                       /mol_type="mRNA"
                       /db_xref="taxon:7227"
                       /clone="GH12341"
       gene            1. .3317
                       /gene="tup"
       CDS             452. .1849
                       /gene="tup"
                       /codon_start=1
                       /product="tailup"
                       /protein_id="AAD38649.1"
                       /db_xref="GI:5052638"
                       /db_xref="FLYBASE:FBcd0012341"
                       /db_xref="FLYBASE:FBgn0003896"
                       /translation="MVMAEIGGHLAHQLPLHNHNHNQTGLQPSLVMNHHLDLDCHGHD
                       VIKKQRLSHCVGCGGQIHDQYILRVAPDLEWHAACLKCQECRQFLDESCTCFVRDGKT
                       YCKRDYVRLFGTKCDKCGNSFSKNDFVMRAKTKIFHIECFRCSACARQLLPGDEFALR
                       DAGALYCKEDHDVLEKSSQSSLTSSSVESNNNISSSNNNNTNLSNNNHSSELGSMSDS
                       GSESGSHKSIRDKRPSGPSDGKPTRVRTVLNEKQLHTLRTCYNANPRPDALMKEQLVE
                       MTSLSPRVIRVWFQNKRCKDKKKTIQMKLQMQQEKEGRKLGYGAMQGIPMIASSPVRH
                       DSPLNLQGLDVQTYQPPWKALSDFALHADLDSNGAINTHTPAFQQLVNQMHGYDLNGM
                       PILPPHPHSHPAQGPPHQHPPPPGGPHNHQNQQPNQQPGGSSLDSGITSHHHPDSTDS
                       YVTYLESDDSMQGSP"
  ORIGIN

  Alignment Scores:
  Pred. No.:                742             Length:         3317
  Score:                    42.00           Matches:        6
  Percent Similarity:       100.00%         Conservative:   0
  Best Local Similarity:    100.00%         Mismatches:     0
  Query Match:              100.00%         Indels:         0
  DB:                       2               Gaps:           0

  US-09-858-332G-28 (1-6) x AF145674 (1-3317)

  Qy          1 HisAsnHisAsnHisAsn 6
                ||||||||||||||||||
  Db        500 CACAATCACAACCACAAT 517

  RESULT 280
  AF111942
  LOCUS       AF111942               3336 bp    DNA     linear   INV 04-MAY-1999
  DEFINITION  Dictyostelium discoideum AX4 development protein DG1106 (DG1106)
              gene, partial cds.
  ACCESSION   AF111942
  VERSION     AF111942.1  GI:4731913
  KEYWORDS    .
  SOURCE      Dictyostelium discoideum
    ORGANISM  Dictyostelium discoideum
              Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
  REFERENCE   1  (bases 1 to 3336)
    AUTHORS   Iranfar,N. and Loomis,W.F.
    TITLE     Dictyostelium discoideum developmental gene DG1106. Disruption of
              this gene results in morphological defect
    JOURNAL   Unpublished
  REFERENCE   2  (bases 1 to 3336)
    AUTHORS   Iranfar,N. and Loomis,W.F.
    TITLE     Direct Submission
    JOURNAL   Submitted (09-DEC-1998) Biology, UCSD, 9500 Gilman Drive, Bonner
              Hall, Room 2321, La Jolla, CA 92093, USA
  FEATURES             Location/Qualifiers
       source          1. .3336
                       /organism="Dictyostelium discoideum"
                       /mol_type="genomic DNA"
                       /strain="AX4"
                       /db_xref="taxon:44689"
       gene            <1. .>2867
                       /gene="DG1106"
                       /note="disruption results in morphological defects"
```

```
     mRNA            <1. .>2867
                     /gene="DG1106"
                     /product="development protein DG1106"
     CDS             <1. .2867
                     /gene="DG1106"
                     /note="similar to Kelch repeats from Drosophila
                     melanogaster ring canal protein presented in GenBank
                     Accession Number L08483"
                     /codon_start=3
                     /product="development protein DG1106"
                     /protein_id="AAD28549.1"
                     /db_xref="GI:4731914"
                     /translation="IHPSSNSGGGGSSQFSPLMVPPLPLNPLSSPSMMQLLTDDETTS
                     VSTLNKYNNNINNNSNHHQHLNHHQPQQPQPQVQPQSQPQPQPQSTFQTQSPIQPPPP
                     IQQYYSKNNNNEQHPSFSNHSPNIFGGFSPLVNGPTNVYNVNVQQSPHIRPTNSSTSS
                     SVPSSPVIMGSNNFNHINNNNNNNINSNNNNNNNNNENINNNHNQFLYSCNNNNNNNT
                     NNINNNNNNNNNNNNIGSNNNNNNNNINNSHHIYKSNSNNNIFINNNSLINNSNNSNN
                     NNITNNNNLIISQTYINDNIFFITNNECLLHNSEIQYYCFECCSGVCAKCIYIDNQHY
                     GHALKSLNDILNETKQDYKATKQYLNQVRNKLHIQREISEQAIQCQPDLNTQKQKIKM
                     KINTITQQLLNQVDQDFNEFKNSLEIKRNLINQLLSDIEFEHKQNKKNEQKQIQQNKR
                     GHYSNNIIINSNNNSNNHQQQIINDQMEYLELGNSNPTSPAQHSSSSSSSITSTSSST
                     SSSTSSISSSSSSSSSSKTSSPNSKGLIQQINSLMNKKKPLLQSIQNSIQVSNIERSS
                     TWCLDLVENKLNLSSSSSSSSSQQQQQQQQQQQQQQQQQQQQQQQQQQQQQNKNNNNNI
                     SAPNSPIQNKLIESNNEEIINIINSTSLNLIRKSNLTIKFTIRFIYAIGGQLKQSVEC
                     YSIQQNKWNYVSPMNFHRNRISAVFDGKQYIYVFGGECPGGFEKSVERYCIHDNKWVM
                     GKEMPRERSRHATVFDGQRYIYIVGGKDQTWFSNQIDRFDILTQEYQSLKKMKVPRSD
                     LSAIYDPRNGGSIYAIGGFNGKALDIIEQYDIQSDKWVKVGKMRKQRDGPGAVYDGNN
                     HIYVMGGSFGGRKISNNLERMNIETWEWESLSDMKKPIDVRNSAIFDGTGNIYVIGGY
                     FSEVLAENQCYNIEKNEWNYLEPMIEPREGNALTYVEFQV"
ORIGIN

Alignment Scores:
Pred. No.:                746             Length:         3336
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       2               Gaps:           0

US-09-858-332G-28 (1-6) x AF111942 (1-3336)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        238 CACAATCACAACCACAAC 255

RESULT 281
AF111943
LOCUS       AF111943               3379 bp    DNA     linear   INV 04-MAY-1999
DEFINITION  Dictyostelium discoideum AX4 development protein DG1124 (DG1124)
            gene, partial cds.
ACCESSION   AF111943
VERSION     AF111943.1  GI:4731915
KEYWORDS    .
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 3379)
  AUTHORS   Iranfar,N. and Loomis,W.F.
  TITLE     Dictyostelium discoideum developmental gene DG1124. Disruption of
            this gene results in morphological defect
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3379)
  AUTHORS   Iranfar,N. and Loomis,W.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1998) Biology, UCSD, 9500 Gilman Drive, Bonner
            Hall, Room 2321, La Jolla, CA 92093, USA
FEATURES             Location/Qualifiers
     source          1. .3379
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
     gene            <1. .>2384
                     /gene="DG1124"
                     /note="disruption results in morphological defects"
     mRNA            join(<1. .368,479. .>2384)
                     /gene="DG1124"
                     /product="development protein DG1124"
     CDS             join(<1. .368,479. .2384)
                     /gene="DG1124"
                     /codon_start=1
                     /product="development protein DG1124"
                     /protein_id="AAD28550.1"
                     /db_xref="GI:4731916"
                     /translation="IDSTGGEEPQQRIEVNQNIPFLIDIKDFGFEKVDIHKTSFSKQQ
                     QQQQQQQQLNKSTTHIKYNSKVMLYSFYSNIEDPSITKKIKFLVSKEDKDISCIGGTW
                     LPDLDGEDVNDHPQTLINTAIRTTKEYTQIDLSGCKKWYKFMEVHYYRPPTEEQSYYQ
                     EITVIYVPDITDIQPMDVDSLLKVPTVSETTPSAPVGVTTATTKTTEEEEEEDQAPKD
                     HVSKDFDNDETSSNTEQNQPPKSSTTSNTTKPAAAVPTTTAATATTTHGSLHLPKASK
                     FYNVITPSSEGNRYKAMMLSLDGLLDYDESDTHEGTFEASLFSELFYLMLCRDMGTNI
                     LTSIVNYRPTSQESDEITDTTLKRKRDDNHAFRGENSEINTSNNTNTNNNNNNNNNNN
                     GTNQPTDSSSLDGNRKSTEKSPQQDVCGVTDKKVKLDGESSTDNTENGEIIGDNTLTN
                     TNNEPVQIEGELNEGLNNNTNNINNNNNNNNEINNNDNNNNNQNENNNQNDNNNEYNI
                     ESNNYNNNDNSSNNTNTSNTSSSSINSPNLSNKDDSNRNSAENQRSSVTSNASSSTST
                     TIGTQVDESAKVLESFHFFDLNRVNYLREEDVETIIHNLGLFLSKSYVQNVVQKVSTN
                     YTSKQGRVYYQEIVKKLIKPTQQQLSTIPSLPQTHSIVPSQSQPQPQQQPQQQEQQQQ
                     QQQQEQQQQQQEQELMSIEQPQQPTQQQDSMAIEQQSQELAIEQESNLDQPPKDQDLV
                     QQEPIDQQIQQNITENI"
ORIGIN

Alignment Scores:
Pred. No.:                756             Length:         3379
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       2               Gaps:           0

US-09-858-332G-28 (1-6) x AF111943 (1-3379)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       2101 CACAATCACAACCACAAC 2118

RESULT 282
AF300334
LOCUS       AF300334               3392 bp    DNA     linear   INV 12-NOV-2000
DEFINITION  Dictyostelium discoideum putative transcription factor StkA (stkA)
            gene, partial cds.
ACCESSION   AF300334
VERSION     AF300334.1  GI:11139097
KEYWORDS    .
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 3392)
  AUTHORS   Ke,C.-H.H., Chen,Y.-H.H., Liu,C.-I.I., Yang,C.-S.S., Gross,J.D.,
            Newell,P.C. and Chang,W.-T.T.
  TITLE     The regulation of cell fate gene stalky expression in Dictyostelium
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3392)
  AUTHORS   Ke,C.-H.H., Yang,C.-S.S., Huang,Y.-C.C. and Chang,W.-T.T.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-AUG-2000) Biochemistry, National Cheng Kung
            University, 1 University Road, Tainan, Taiwan 70101, Republic of
            China
FEATURES             Location/Qualifiers
     source          1. .3392
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="Ax2"
                     /db_xref="taxon:44689"
                     /note="44689"
     gene            1. .>3392
                     /gene="stkA"
     promoter        1. .2897
```

```
                /gene="stkA"
     mRNA       2898. .>3392
                /gene="stkA"
                /product="putative transcription factor StkA"
     CDS        3390. .>3392
                /gene="stkA"
                /note="knock-out prevents ability to complete spore
                formation, leading to the formation of a fruiting body
                containing only stalk cells"
                /codon_start=1
                /product="putative transcription factor StkA"
                /protein_id="AAG31607.1"
                /db_xref="GI:11139098"
                /translation="M"
ORIGIN
```

Alignment Scores:

Pred. No.:	758	Length:	3392
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-28 (1-6) x AF300334 (1-3392)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        725 CACAACCACAACCACAAC 742
```

```
RESULT 283
AR146444
LOCUS       AR146444                3455 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 3 from patent US 6218512.
ACCESSION   AR146444
VERSION     AR146444.1  GI:15109633
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3455)
  AUTHORS   Guarente,L.P., Austriaco,N. Jr. and Kennedy,B.
  TITLE     Genes determining cellular senescence in yeast
  JOURNAL   Patent: US 6218512-A 3 17-APR-2001;
FEATURES             Location/Qualifiers
     source          1. .3455
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN
```

Alignment Scores:

Pred. No.:	772	Length:	3455
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-09-858-332G-28 (1-6) x AR146444 (1-3455)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       2487 CATAATCATAATCATAAT 2504
```

```
RESULT 284
AR579950
LOCUS       AR579950                3455 bp    DNA     linear   PAT 15-DEC-2004
DEFINITION  Sequence 3 from patent US 6787300.
ACCESSION   AR579950
VERSION     AR579950.1  GI:56609846
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3455)
  AUTHORS   Guarente,L.P., Austriaco,N. Jr. and Kennedy,B.
  TITLE     Identifying lifespan-altering agents
  JOURNAL   Patent: US 6787300-A 3 07-SEP-2004;
            Massachusetts Institute of Technology; Cambridge, MA
FEATURES             Location/Qualifiers
     source          1. .3455
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN
```

Alignment Scores:

Pred. No.:	772	Length:	3455
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-09-858-332G-28 (1-6) x AR579950 (1-3455)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       2487 CATAATCATAATCATAAT 2504
```

```
RESULT 285
YSCMPT5
LOCUS       YSCMPT5                 3455 bp    DNA     linear   PLN 06-FEB-1999
DEFINITION  Yeast MPT5 gene for suppressor protein, complete cds.
ACCESSION   D26184
VERSION     D26184.1  GI:450489
KEYWORDS    multicopy suppressor; suppressor protein.
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 3455)
  AUTHORS   Sakai,A.
  TITLE     Multicopy suppressors of the yeast pop2 mutation
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3455)
  AUTHORS   Sakai,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-DEC-1993) Akira Sakai, Mitsubishi Kasei Institute of
            Life Sciences, Molecular Biology; 11, Minamiooya, Machida, Tokyo
            194, Japan (Tel:0427-24-6258, Fax:0427-29-1252)
COMMENT     Submitted (23-Dec-1993) to DDBJ by:
            Akira Sakai
            Department of Molecular Genetics
            Mitsubishi Kasei Institute of Life Sicneces
            11 Minamiooya, Machida-shi
            Tokyo 194
            Japan
            Phone:  0427-24-6258
            Fax:    0427-24-6316.
FEATURES             Location/Qualifiers
     source          1. .3455
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /strain="S288C"
                     /db_xref="taxon:4932"
     gene            663. .3167
                     /gene="MPT5"
     CDS             663. .3167
                     /gene="MPT5"
                     /note="multicopy suppressor of pop2 mutation"
                     /codon_start=1
                     /product="suppressor protein"
                     /protein_id="BAA05172.1"
                     /db_xref="GI:450490"
                     /translation="MSYNHQPQLSINSVQSLLEPVTPPPLGQMNNKRNHQKAHSLDLS
                     GFNQFISSTQSPLALMNNTSTSNSANSFSPNPNAASNSTGLSASMANPPAILPLINEF
```

DLEMDGPRRKSSHDFTVVAPSNSGVNTSSLIMETPSSSVTPAASLRNFSNSNNAASKC
GVDNSSFGLSSSTSSSMVEISALPLRDLDYIKLATDQFGCRFLQKKLETPSESNMVRD
LMYEQIKPFFLDLILDPFGNYLVQKLCDYLTAEQKTLLIQTIYPNVFQISINQYGTRS
LQKIIDTVDNEVQIDLIIKGFSQEFTSIEQVVTLINDLNGNHVIQKCIFKFSPSKFGF
IIDAIVEQNNIITISTHKHGCCVLQKLLSVCTLQQIFKISVKIVQFLPGLINDQFGNY
IIQFLLDIKELDFYLLAELFNRLSNELCQLSCLKFSSNVVEKFIKKLFRIITGFIVNN
NGGASQRTAVASDDVINASMNILLTTIDIFTVNLNVLIRDNFGNYALQTLLDVKNYSP
LLAYNKNSNAIGQNSSSTLNYGNFCNDFSLKIGNLIVLTKELLPSIKTTSYAKKIKLK
VKAYAEATGIPFTDISPQVTAMSHNNLQTINNENKNPHNKNSHNHNHNHNHNHAHNNN
NNNNQKSHTRHFSLPANAYHRRSNSSVTNNFSNQYAQDQKIHSPQQIMNFNQNAYPSM
GAPSFNSQTNPPLVSHNSLQNFDNRQFANLMAHPNSAAPIHSFSSSNITNVNPNVSRG
FKQPGFMMNETDKINANHFSPYSNANSQNFNESFVPRMQYQTEGANWDSSLSMKSQHI
GQGPYNQVNMSRNASISNMPAMNTARTSDELQFTLP"

ORIGIN

Alignment Scores:

Pred. No.:	772	Length:	3455
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	15	Gaps:	0

US-09-858-332G-28 (1-6) x YSCMPT5 (1-3455)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       2487 CATAATCATAATCATAAT 2504
```

RESULT 286
CS082046/c
LOCUS CS082046 3501 bp DNA linear PAT 18-MAY-2005
DEFINITION Sequence 16 from Patent WO2005040421.
ACCESSION CS082046
VERSION CS082046.1 GI:66349024
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1
AUTHORS Widschwendter,M.
TITLE Prognostic and diagnostic markers for cell proliferative disorders of the breast tissues
JOURNAL Patent: WO 2005040421-A 16 06-MAY-2005;
Widschwendter, Martin (AT)
FEATURES Location/Qualifiers
source 1. .3501
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

Alignment Scores:

Pred. No.:	782	Length:	3501
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-09-858-332G-28 (1-6) x CS082046 (1-3501)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1478 CACAACCACAACCACAAC 1461
```

RESULT 287
SCYLR055C/c
LOCUS SCYLR055C 3605 bp DNA linear PLN 11-AUG-1997
DEFINITION S.cerevisiae chromosome XII reading frame ORF YLR055c.
ACCESSION Z73227 Y13138
VERSION Z73227.1 GI:1360395
KEYWORDS .
SOURCE Saccharomyces cerevisiae (baker's yeast)
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 3605)
AUTHORS Andre,B. and Urrestarazu,L.A.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3605)
AUTHORS MIPS.
TITLE Direct Submission
JOURNAL Submitted (22-MAY-1996) Data collected by MIPS on behalf of the European yeast chromosome XII sequencing project. MIPS at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152 Martinsried, FRG; E-mail: Mewes@mips.embnet.org
FEATURES Location/Qualifiers
source 1. .3605
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/db_xref="taxon:4932"
/chromosome="XII"
gene complement(1017. .2825)
/gene="SPT8"
CDS complement(1017. .2825)
/gene="SPT8"
/note="ORF YLR055c"
/codon_start=1
/protein_id="CAA97585.1"
/db_xref="GI:1360396"
/db_xref="GOA:P38915"
/db_xref="SGD:S0004045"
/db_xref="UniProt/Swiss-Prot:P38915"
/translation="MDEVDDILINNQVVDDEEDDEEMLSGLENDSKQDLEGNDDGGED
EEDDDDDDEDDDDDEDERDDDEQEDDDGEDDAARMDKTATPTNEHQHDEQKAAAAGA
GGAGDSGDAVTKIGSEDVKLSDVDGGVGSREASSSTHEASANGEVYEYYKHMLNAAQI
ADSYNIYPTAAIPIQTHVNALAVSRGLKYLFLGGSDGYIRKYDLLNTLEGKLSLTILQ
KHSLAESIQNAGILQSYWENEIPQKKSEMKLSANKTDYEPKVSPVHSLEVQSECLFIL
SGLQNGGITMQGVRYMEGSIAHYFKGRNGHTQIVNILRLNGQEDRFLSGSWDKRLLEW
DLQTGDIVNEFKKSRSELSSLEMRPLYSSVDVSGNVNSGKENENADDDMDSLFGDEDE
DEKQDAGNEPVETGDGSNGEENKEQISEESLNIVYDESVFMTSGLNGSVHIWDRRMTQ
SPALSLERGAGVPPWCLSACWGVDGDHVYAGRRNACVEQFDLKMPSKPIHNLKLPSIS
GPVSCVKAMPNNKHLLCASRDNIRLYNVEIAVDASNSTTKSSKVPFLIVPGHHGGIIS
NLYLDPTSRFIISTSGNRGWQGNSTDTTLIYDIDLE"

ORIGIN

Alignment Scores:

Pred. No.:	804	Length:	3605
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	15	Gaps:	0

US-09-858-332G-28 (1-6) x SCYLR055C (1-3605)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        999 CATAACCATAATCATAAT 982
```

RESULT 288
YSCHTR1
LOCUS YSCHTR1 3769 bp DNA linear PLN 29-MAY-2002
DEFINITION Saccharomyces cerevisiae HTR1 gene, complete cds.
ACCESSION D25541
VERSION D25541.1 GI:436256
KEYWORDS .
SOURCE Saccharomyces cerevisiae (baker's yeast)
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 3769)
AUTHORS Kikuchi,Y., Oka,Y., Kobayashi,M., Uesono,Y., Toh-e,A. and Kikuchi,A.

```
  TITLE     A new yeast gene, HTR1, required for growth at high temperature, is
            needed for recovery from mating pheromone-induced G1 arrest
  JOURNAL   Mol. Gen. Genet. 245 (1), 107-116 (1994)
   PUBMED   7845352
REFERENCE   2  (bases 1 to 3769)
  AUTHORS   Kikuchi,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-NOV-1993) Akihiko Kikuchi, Mitsubishi Kasei Institute
            of Life Sciences; 11 Minamiooya, Machida, Tokyo 194, Japan
            (Tel:0427-24-6229, Fax:0427-24-6317)
COMMENT     Submitted (26-Nov-1993) to DDBJ by:
            Akihiko Kikuchi
            Mitsubishi Kasei Institute of Life Sciences
            11 Minamiooya, Machida-shi
            Tokyo 194
            Japan
            Phone: 0427-24-6229
            Fax:   0427-24-6317.
FEATURES             Location/Qualifiers
     source          1. .3769
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /strain="ATCC70391"
                     /db_xref="taxon:4932"
                     /chromosome="12"
                     /map="12 L"
     gene            665. .3169
                     /gene="HTR1"
     CDS             665. .3169
                     /gene="HTR1"
                     /codon_start=1
                     /protein_id="BAA05024.1"
                     /db_xref="GI:469512"
                     /translation="MSYNHQPQLSINSVQSLLEPVTPPPLGQMNNKRNHQKAHSLDLS
                     GFNQFISSTQSPLALMNNTSTSNSANSFSPNPNAASNSTGLSASMANPPAILPLINEF
                     DLEMDGPRRKSSHDFTVVAPSNSGVNTSSLIMETPSSSVTPAASLRNFSNSNNAASKC
                     GVDNSSFGLSSSTSSSMVEISALPLRDLDYIKLATDQFGCRFLQKKLETPSESNMVRD
                     LMYEQIKPFFLDLILDPFGNYLVQKLCDYLTAEQKTLLIQTIYPNVFQISINQYGTRS
                     LQKIIDTVDNEVQIDLIIKGFSQEFTSIEQVVTLINDLNGNHVIQKCIFKFSPSKFGF
                     IIDAIVEQNNIITISTHKHGCCVLQKLLSVCTLQQIFKISVKIVQFLPGLINDQFGNY
                     IIQFLLDIKELDFYLLAELFNRLSNELCQLSCLKFSSNVVEKFIKKLFRIITGFIVNN
                     NGGASQRTAVASDDVINASMNILLTTIDIFTVNLNVLIRDNFGNYALQTLLDVKNYSP
                     LLAYNKNSNAIGQNSSSTLNYGNFCNDFSLKIGNLIVLTKELLPSIKTTSYAKKIKLK
                     VKAYAEATGIPFTDISPQVTAMSHNNLQTINNENKNFHNKNSHNHNHNHNHNHAHNNN
                     NNNNQKSHTRHFSLPANAYHRRSNSSVTNNFSNQYAQDQKIHSPQQIMNFNQNAYPSM
                     GAPSFNSQTNPPLVSHNSLQNFDNRQFANLMAHPNSAAPIHSFSSSNITNVNPNVSRG
                     FKQPGFMMNETDKINANHFSPYSNANSQNFNESFVPRMQYQTEGANWDSSLSMKSQHI
                     GQGPYNQVNMSRNASISNMPAMNTARTSDELQFTLP"
ORIGIN

Alignment Scores:
Pred. No.:                839             Length:         3769
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       15              Gaps:           0

US-09-858-332G-28 (1-6) x YSCHTR1 (1-3769)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       2489 CATAATCATAATCATAAT 2506

RESULT 289
CAALK8
LOCUS       CAALK8                  3815 bp    DNA     linear   PLN 18-APR-2005
DEFINITION  Candida albicans alk8 gene.
ACCESSION   Y14766
VERSION     Y14766.1  GI:3395457
KEYWORDS    alk8 gene; multidrug resistance.
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1
  AUTHORS   Prasad,R.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3815)
  AUTHORS   Prasad,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-1997) R. Prasad, Jawaharlal Nehru University,
            School of Life Sciences, New Delhi 110 167, INDIA
  REMARK    revised by author 06-NOV-97, 13-JUL-98, 20-JUL-98
COMMENT     Related sequences: D12719, Z13010 and X55881.
FEATURES             Location/Qualifiers
     source          1. .3815
                     /organism="Candida albicans"
                     /mol_type="genomic DNA"
                     /strain="1001"
                     /db_xref="taxon:5476"
                     /clone="pNC36"
                     /dev_stage="bud"
                     /note="fosmids, 1D1, 2D2, 2H8, 8A7, 15H3, 19A1, 19A2,
                     17D5"
     gene            1943. .3490
                     /gene="alk8"
     CDS             1943. .3490
                     /gene="alk8"
                     /function="multidrug resistance gene"
                     /codon_start=1
                     /transl_table=12
                     /protein_id="CAA75058.1"
                     /db_xref="GI:3395458"
                     /db_xref="GOA:O74626"
                     /db_xref="InterPro:IPR001128"
                     /db_xref="InterPro:IPR002401"
                     /db_xref="InterPro:IPR002974"
                     /db_xref="UniProt/TrEMBL:O74626"
                     /translation="MLEEISFESWYLYLIICGVLYFCTNYLYIQYLNKKLGAKSPGSC
                     KSDGCFGFRSGFEMLKAKKIGRQVDLLHERFNDLSNPEVNDTFTSYNFGFPMIITKDP
                     ENIKALLATQFNDFSLGRRYDFFKPLLGKGIFTLDGEGWKHSRSMLRPQFAREQIAHV
                     KALEPHFQLLKKHITKNKGQFFDIQELFFRFTVDSATEFLFGESVSSLKDESIGYNQD
                     EFDFAGRKDFFPEAFNKSQIYLSTRSLLQNLYWLVNPKDFQNCNAIVHKFSDYYVNKAL
                     NATPEELEKNSGYVFLYELVKQTRDPQVLRDQALNILLAGRDTTAGLLSFAVFELARN
                     PQVWKKLREEIIERYADGVTEISFETLKQCEYLKAVINETLRLYPSVPRNGRFANKNT
                     TLPRGGGPDGMSTILIKQGQAVMYSIYSTQRDPKYYGKDADVFRPERWFEPETRKLGW
                     AFLPFNGGPRICLGQQFALTEASYVLVRLAQCCENLQLTEGIEYPPAKLTHLTMCMFD
                     GVPIKVD"
ORIGIN

Alignment Scores:
Pred. No.:                849             Length:         3815
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       15              Gaps:           0

US-09-858-332G-28 (1-6) x CAALK8 (1-3815)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        895 CACAACCACAACCACAAC 912

RESULT 290
AF332954
LOCUS       AF332954                3897 bp    DNA     linear   PLN 03-JUN-2005
DEFINITION  Zea mays transposon Ds2(D4), complete sequence.
ACCESSION   AF332954
VERSION     AF332954.1  GI:13345975
KEYWORDS    .
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
```

```
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 3897)
  AUTHORS   Yan,X., Martinez-Perez,I.M., Kavchok,S. and Dooner,H.K.
  TITLE     Origination of Ds elements from Ac elements in maize: evidence for
            rare repair synthesis at the site of Ac excision
  JOURNAL   Genetics 152 (4), 1733-1740 (1999)
   PUBMED   10430597
REFERENCE   2  (bases 1 to 3897)
  AUTHORS   Yan,X., Martinez-Perez,I.M., Kavchok,S. and Dooner,H.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JAN-2001) Waksman Institute, Rutgers University,
            Piscataway, NJ 08855, USA
FEATURES             Location/Qualifiers
     source          1. .3897
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4577"
                     /chromosome="9"
                     /map="9S"
                     /note="internal deletion derivative 4 of Ac element in
                     bz-m2(Ac)"
     repeat_region   1. .3897
                     /note="internal deletion derivative 1 of Ac element in
                     bz-m2(Ac)"
                     /transposon="Ds2(D4)"
ORIGIN

Alignment Scores:
Pred. No.:              867           Length:        3897
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     15            Gaps:          0

US-09-858-332G-28 (1-6) x AF332954 (1-3897)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1420 CACAACCACAACCACAAC 1437

RESULT 291
CQ582980/c
LOCUS       CQ582980               3929 bp    DNA     linear   PAT 02-FEB-2004
DEFINITION  Sequence 10738 from Patent WO0171042.
ACCESSION   CQ582980
VERSION     CQ582980.1  GI:41643966
KEYWORDS    .
SOURCE      Drosophila sp.
  ORGANISM  Drosophila sp.
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1
  AUTHORS   Venter,J.C., Adams,M., Li,P.W. and Myers,E.W.
  TITLE     Detection kits, such as nucleic acid arrays, for detecting the
            expression of 10,000 or more Drosophila genes and uses thereof
  JOURNAL   Patent: WO 0171042-A 10738 27-SEP-2001;
            PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1. .3929
                     /organism="Drosophila sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:7242"
ORIGIN

Alignment Scores:
Pred. No.:              874           Length:        3929
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-28 (1-6) x CQ582980 (1-3929)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1717 CACAACCACAACCACAAC 1700

RESULT 292
PUFFC4PR/c
LOCUS       PUFFC4PR               3990 bp    DNA     linear   INV 16-MAR-2001
DEFINITION  Bradysia hygida PC4-1 (PC4-1) gene, promoter, partial cds.
ACCESSION   U13892
VERSION     U13892.2  GI:13366262
KEYWORDS    .
SOURCE      Bradysia hygida
  ORGANISM  Bradysia hygida
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Sciaroidea;
            Sciaridae; Bradysia.
REFERENCE   1  (bases 3485 to 3990)
  AUTHORS   Monesi,N., Fernandez,M.A., Fontes,A.M., Basso,L.R. Jr.,
            Nakanishi,Y., Baron,B., Buttin,G. and Paco-Larson,M.L.
  TITLE     Molecular characterization of an 18 kb segment of DNA puff C4 of
            Bradysia hygida (Diptera, sciaridae)
  JOURNAL   Chromosoma 103 (10), 715-724 (1995)
   PUBMED   7664619
REFERENCE   2  (bases 1 to 3990)
  AUTHORS   Paco-Larson,M.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-AUG-1994) Maria Luisa M. Paco-Larson, Morfologia,
            Faculdade de Medicina de Ribeirao Preto, Universidade de Sao Paulo,
            Av. Bandeirantes 3900, Ribeirao Preto, Sao Paulo, Brasil, 14049-900
REFERENCE   3  (bases 1 to 3990)
  AUTHORS   Basso,L.R. Jr. and Paco-Larson,M.L.
  TITLE     BhC4-1 promoter region
  JOURNAL   Unpublished
COMMENT     On Mar 16, 2001 this sequence version replaced gi:836886.
FEATURES             Location/Qualifiers
     source          1. .3990
                     /organism="Bradysia hygida"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:35572"
                     /chromosome="C"
                     /map="4"
                     /clone="pBC40.3E/H,pBC40.8H"
                     /sex="female"
                     /tissue_type="whole animals"
                     /dev_stage="adult"
     promoter        1. .3696
     TATA_signal     3669. .3674
     gene            3736. .3990
                     /gene="PC4-1"
     CDS             join(3736. .3780,3848. .>3990)
                     /gene="PC4-1"
                     /codon_start=1
                     /protein_id="AAA83555.2"
                     /db_xref="GI:13366263"
                     /translation="MKCFLILVIAVVALQCVSANPSRKRDCDCDCDCDKRRRIDDDCD
                     DYVSTLMMTIVLTITKLNS"
ORIGIN

Alignment Scores:
Pred. No.:              887           Length:        3990
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     2             Gaps:          0

US-09-858-332G-28 (1-6) x PUFFC4PR (1-3990)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
```

```
Db       3896 CACAATCACAATCACAAT 3879

RESULT 293
CQ572621/c
LOCUS       CQ572621               4048 bp    DNA     linear   PAT 02-FEB-2004
DEFINITION  Sequence 379 from Patent WO0171042.
ACCESSION   CQ572621
VERSION     CQ572621.1  GI:41637060
KEYWORDS    .
SOURCE      Drosophila sp.
  ORGANISM  Drosophila sp.
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1
  AUTHORS   Venter,J.C., Adams,M., Li,P.W. and Myers,E.W.
  TITLE     Detection kits, such as nucleic acid arrays, for detecting the
            expression of 10,000 or more Drosophila genes and uses thereof
  JOURNAL   Patent: WO 0171042-A 379 27-SEP-2001;
            PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1. .4048
                     /organism="Drosophila sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:7242"
ORIGIN

Alignment Scores:
Pred. No.:              899           Length:         4048
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     6             Gaps:           0

US-09-858-332G-28 (1-6) x CQ572621 (1-4048)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       3733 CATAATCATAATCATAAT 3716

RESULT 294
CQ589985/c
LOCUS       CQ589985               4054 bp    DNA     linear   PAT 02-FEB-2004
DEFINITION  Sequence 17743 from Patent WO0171042.
ACCESSION   CQ589985
VERSION     CQ589985.1  GI:41648636
KEYWORDS    .
SOURCE      Drosophila sp.
  ORGANISM  Drosophila sp.
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1
  AUTHORS   Venter,J.C., Adams,M., Li,P.W. and Myers,E.W.
  TITLE     Detection kits, such as nucleic acid arrays, for detecting the
            expression of 10,000 or more Drosophila genes and uses thereof
  JOURNAL   Patent: WO 0171042-A 17743 27-SEP-2001;
            PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1. .4054
                     /organism="Drosophila sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:7242"
ORIGIN

Alignment Scores:
Pred. No.:              901           Length:         4054
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     6             Gaps:           0

US-09-858-332G-28 (1-6) x CQ589985 (1-4054)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        338 CATAACCATAACCATAAC 321

RESULT 295
AX489484
LOCUS       AX489484               4146 bp    DNA     linear   PAT 16-AUG-2002
DEFINITION  Sequence 6784 from Patent WO02053728.
ACCESSION   AX489484
VERSION     AX489484.1  GI:22323496
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1
  AUTHORS   Roemer,T., Jiang,B., Boone,C., Bussey,H. and Ohlsen,K.L.
  TITLE     Gene disruption methodologies for drug target discovery
  JOURNAL   Patent: WO 02053728-A 6784 11-JUL-2002;
            Elitra Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .4146
                     /organism="Candida albicans"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:5476"
ORIGIN

Alignment Scores:
Pred. No.:              920           Length:         4146
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     6             Gaps:           0

US-09-858-332G-28 (1-6) x AX489484 (1-4146)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        170 CACAACCACAACCACAAC 187

RESULT 296
AX536824
LOCUS       AX536824               4146 bp    DNA     linear   PAT 22-NOV-2002
DEFINITION  Sequence 425 from Patent WO02064766.
ACCESSION   AX536824
VERSION     AX536824.1  GI:25263259
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1
  AUTHORS   Contreras,R.H., Eberhardt,I., Luyten,W.H. and Reekmans,R.J.
  TITLE     Bax-responsive genes for drug target identification in yeast and
            fungi
  JOURNAL   Patent: WO 02064766-A 425 22-AUG-2002;
            JANSSEN PHARMACEUTICA N.V. (BE)
FEATURES             Location/Qualifiers
     source          1. .4146
                     /organism="Candida albicans"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:5476"
ORIGIN

Alignment Scores:
Pred. No.:              920           Length:         4146
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
```

```
Query Match:            100.00%        Indels:        0
DB:                     6              Gaps:          0

US-09-858-332G-28 (1-6) x AX536824 (1-4146)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        170 CACAACCACAACCACAAC 187

RESULT 297
AF332955
LOCUS       AF332955                4167 bp    DNA     linear   PLN 03-JUN-2005
DEFINITION  Zea mays transposon Ds2(D5), complete sequence.
ACCESSION   AF332955
VERSION     AF332955.1  GI:13345976
KEYWORDS    .
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 4167)
  AUTHORS   Yan,X., Martinez-Ferez,I.M., Kavchok,S. and Dooner,H.K.
  TITLE     Origination of Ds elements from Ac elements in maize: evidence for
            rare repair synthesis at the site of Ac excision
  JOURNAL   Genetics 152 (4), 1733-1740 (1999)
   PUBMED   10430597
REFERENCE   2  (bases 1 to 4167)
  AUTHORS   Yan,X., Martinez-Ferez,I.M., Kavchok,S. and Dooner,H.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JAN-2001) Waksman Institute, Rutgers University,
            Piscataway, NJ 08855, USA
FEATURES             Location/Qualifiers
     source          1. .4167
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4577"
                     /chromosome="9"
                     /map="9S"
                     /note="internal deletion derivative 5 of Ac element in
                     bz-m2(Ac)"
     repeat_region   1. .4167
                     /note="internal deletion derivative 1 of Ac element in
                     bz-m2(Ac)"
                     /transposon="Ds2(D5)"
ORIGIN

Alignment Scores:
Pred. No.:              925            Length:        4167
Score:                  42.00          Matches:       6
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     15             Gaps:          0

US-09-858-332G-28 (1-6) x AF332955 (1-4167)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1420 CACAACCACAACCACAAC 1437

RESULT 298
CQ608255
LOCUS       CQ608255                4229 bp    DNA     linear   PAT 02-FEB-2004
DEFINITION  Sequence 36013 from Patent WO0171042.
ACCESSION   CQ608255
VERSION     CQ608255.1  GI:41660816
KEYWORDS    .
SOURCE      Drosophila sp.
  ORGANISM  Drosophila sp.
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1
  AUTHORS   Venter,J.C., Adams,M., Li,P.W. and Myers,E.W.
  TITLE     Detection kits, such as nucleic acid arrays, for detecting the
            expression of 10,000 or more Drosophila genes and uses thereof
  JOURNAL   Patent: WO 0171042-A 36013 27-SEP-2001;
            PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1. .4229
                     /organism="Drosophila sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:7242"
ORIGIN

Alignment Scores:
Pred. No.:              938            Length:        4229
Score:                  42.00          Matches:       6
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     6              Gaps:          0

US-09-858-332G-28 (1-6) x CQ608255 (1-4229)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1201 CACAACCATAATCACAAT 1218

RESULT 299
CQ578165
LOCUS       CQ578165                4272 bp    DNA     linear   PAT 02-FEB-2004
DEFINITION  Sequence 5923 from Patent WO0171042.
ACCESSION   CQ578165
VERSION     CQ578165.1  GI:41640756
KEYWORDS    .
SOURCE      Drosophila sp.
  ORGANISM  Drosophila sp.
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1
  AUTHORS   Venter,J.C., Adams,M., Li,P.W. and Myers,E.W.
  TITLE     Detection kits, such as nucleic acid arrays, for detecting the
            expression of 10,000 or more Drosophila genes and uses thereof
  JOURNAL   Patent: WO 0171042-A 5923 27-SEP-2001;
            PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1. .4272
                     /organism="Drosophila sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:7242"
ORIGIN

Alignment Scores:
Pred. No.:              947            Length:        4272
Score:                  42.00          Matches:       6
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     6              Gaps:          0

US-09-858-332G-28 (1-6) x CQ578165 (1-4272)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       2914 CATAATCACAACCATAAT 2931

RESULT 300
AR508754/c
LOCUS       AR508754                4285 bp    DNA     linear   PAT 22-SEP-2004
DEFINITION  Sequence 13714 from patent US 6703491.
ACCESSION   AR508754
VERSION     AR508754.1  GI:52444229
KEYWORDS    .
```

```
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 4285)
  AUTHORS   Homburger,S.A., Ebens,A.J. Jr., Erickson,C.S., Francis-Lang,H.L.,
            Margolis,J.S., Reddy,B.P., Ruddy,D.A. and Buchman,A.R.
  TITLE     Drosophila sequences
  JOURNAL   Patent: US 6703491-A 13714 09-MAR-2004;
            Exelixis, Inc.; South San Francisco, CA
FEATURES             Location/Qualifiers
     source          1. .4285
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:                950             Length:         4285
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x AR508754 (1-4285)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       3330 CATAATCATAATCATAAT 3313

RESULT 301
CT025243/c
LOCUS       CT025243                4296 bp    mRNA    linear   VRT 29-JUL-2005
DEFINITION  Xenopus tropicalis finished cDNA, clone TEgg085b11.
ACCESSION   CT025243
VERSION     CT025243.1  GI:71532627
KEYWORDS    .
SOURCE      Xenopus tropicalis (Silurana tropicalis)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 4296)
  AUTHORS   Amaya,E., Ashurst,J.L., Bonfield,J.K., Croning,M.D.R., Davies,R.M.,
            Francis,M.D., Garrett,N., Gilchrist,M.J., Grafham,D.V.,
            McLaren,S.R., Papalopulu,N., Rogers,J., Smith,J.C., Taylor,R.G.,
            Voigt,J. and Zorn,A.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-2005) Sanger Institute, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: trop@sanger.ac.uk
COMMENT     Sanger Xenopus tropicalis EST/cDNA project.
            This sequence is from a Xenopus Gene Collection (XGC) library, from
            a library constructed by Aaron M. Zorn. cDNA was prepared from RNA
            extracted from eggs. EcoRI-NotI cut cDNA was then ligated into
            pCS107 with EcoRI at the 5' end and NotI at the 3' end.
            Vector: pCS107; Site_1: EcoRI; Site_2: NotI
            Host: Escherichia coli XL1-blue.
FEATURES             Location/Qualifiers
     source          1. .4296
                     /organism="Xenopus tropicalis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8364"
                     /clone="TEgg085b11"
                     /clone_lib="XGC-egg"
                     /dev_stage="egg"
ORIGIN

Alignment Scores:
Pred. No.:                952             Length:         4296
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       5               Gaps:           0

US-09-858-332G-28 (1-6) x CT025243 (1-4296)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1300 CATAATCATAATCATAAT 1283

RESULT 302
CS124702/c
LOCUS       CS124702                4492 bp    DNA     linear   PAT 21-JUL-2005
DEFINITION  Sequence 388 from Patent WO2005059172.
ACCESSION   CS124702
VERSION     CS124702.1  GI:71057857
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Foekens,J.
  TITLE     Method and nucleic acids for the improved treatment of breast cell
            proliferative disorders
  JOURNAL   Patent: WO 2005059172-A 388 30-JUN-2005;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .4492
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

Alignment Scores:
Pred. No.:                994             Length:         4492
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x CS124702 (1-4492)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        258 CACAACCATAACCACAAC 241

RESULT 303
DD013389/c
LOCUS       DD013389                4565 bp    DNA     linear   PAT 02-FEB-2005
DEFINITION  Methods and Compositions to Reduce or Eliminate Transmission of a
            Transgene.
ACCESSION   DD013389
VERSION     DD013389.1  GI:58530792
KEYWORDS    JP 2004512007-A/3.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 4565)
  AUTHORS   Dellaporta,S.L. and Moreno,M.A.
  TITLE     Methods and Compositions to Reduce or Eliminate Transmission of a
            Transgene
  JOURNAL   Patent: JP 2004512007-A 3 22-APR-2004;
            Stephen L Dellaporta,Maria A Moreno,Yale University
COMMENT     OS   Zea mays
            PN   JP 2004512007-A/3
            PD   22-APR-2004
            PF   28-FEB-2001 JP 2001563615
            PR   28-FEB-2000 US    60/185524
            PI   stephen l dellaporta,maria a moreno
            CC   Transposable element Ac
            FH   Key                Location/Qualifiers.
FEATURES             Location/Qualifiers
```

```
     source          1. .4565
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4577"
ORIGIN

Alignment Scores:
Pred. No.:              1.01e+03        Length:         4565
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x DD013389 (1-4565)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       3146 CACAACCACAACCACAAC 3129

RESULT 304
AR542626/c
LOCUS       AR542626                4565 bp    DNA     linear   PAT 08-OCT-2004
DEFINITION  Sequence 3 from patent US 6743968.
ACCESSION   AR542626
VERSION     AR542626.1  GI:53935114
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 4565)
  AUTHORS   Dellaporta,S.L. and Moreno,M.A.
  TITLE     Methods and compositions to reduce or eliminate transmission of a
            transgene
  JOURNAL   Patent: US 6743968-A 3 01-JUN-2004;
            Yale University; New Haven, CT
FEATURES             Location/Qualifiers
     source          1. .4565
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              1.01e+03        Length:         4565
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x AR542626 (1-4565)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       3146 CACAACCACAACCACAAC 3129

RESULT 305
AX235897/c
LOCUS       AX235897                4565 bp    DNA     linear   PAT 26-SEP-2001
DEFINITION  Sequence 3 from Patent WO0164926.
ACCESSION   AX235897
VERSION     AX235897.1  GI:15795786
KEYWORDS    .
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1
  AUTHORS   Dellaporta,S.L. and Moreno,M.A.
  TITLE     Methods and compositions to reduce or eliminate transmission of a
            transgene
  JOURNAL   Patent: WO 0164926-A 3 07-SEP-2001;
            YALE UNIVERSITY (US)
FEATURES             Location/Qualifiers
     source          1. .4565
                     /organism="Zea mays"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:4577"
                     /note="Transposable element Ac"
ORIGIN

Alignment Scores:
Pred. No.:              1.01e+03        Length:         4565
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x AX235897 (1-4565)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       3146 CACAACCACAACCACAAC 3129

RESULT 306
ZMAC1
LOCUS       ZMAC1                   4565 bp    mRNA    linear   PLN 10-FEB-1999
DEFINITION  Maize transposable element Activator (Ac) major transcript.
ACCESSION   X05424 X05425
VERSION     X05424.1  GI:22112
KEYWORDS    Ac element; transposable element.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1
  AUTHORS   Kunze,R., Stochaj,U., Laufs,J. and Starlinger,P.
  TITLE     Transcription of transposable element Activator (Ac) of Zea mays L
  JOURNAL   EMBO J. 6, 1555-1563 (1987)
REFERENCE   2  (bases 1 to 4565)
  AUTHORS   Kunze,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-FEB-1988) R. Kunze, Institut fuer Genetik,
            Universitaet zu Koeln, D-5000 Koeln 41, West Germany
FEATURES             Location/Qualifiers
     source          1. .4565
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="taxon:4577"
     misc_feature    303
                     /note="minor transcription start site"
     misc_feature    334
                     /note="major transcription start site"
     misc_feature    343
                     /note="minor transcription start site"
     misc_feature    359
                     /note="minor transcription start site"
     misc_feature    363
                     /note="minor transcription start site"
     misc_feature    569. .2235
                     /note="pcAcP"
     CDS             join(988. .1111,1219. .2566,2638. .3434,3524. .3615,
                     4003. .4065)
                     /codon_start=1
                     /product="ORFa"
                     /protein_id="CAA29005.1"
                     /db_xref="GI:22113"
                     /db_xref="GOA:P08770"
                     /db_xref="UniProt/Swiss-Prot:P08770"
                     /translation="MTPPVGNNPPSGSAIRLAKLMSTTRAPSTRKTNSVFSAYAQGLK
                     RKAEASSSRIQNVRARARGHGCGRTSPSSSTAEAERHFIQSVSSSNANGTATDPSQDD
                     MAIVHEPQPQPQPQPEPQPQPQPEPEEEAPQKRAKKCTSDVWQHFTKKEIEVEVDGKK
                     YVQVWGHCNFPNCKAKYRAEGHHGTSGFRNHLRTSHSLVKGQLCLKSEKDHGKDINLI
```

```
                     EPYKYDEVVSLKKLHLAIIMHEYPFNIVEHEYFVEFVKSLRPHFPIKSRVTARKYIMD
                     LYLEEKEKLYGKLKDVQSRFSTTMDMWTSCQNKSYMCVTIHWIDDDWCLQKRIVGFFH
                     VEGRHTGQRLSQTFTAIMVKWNIEKKLFALSLDNASANEVAVHDIIEDLQDTDSNLVC
                     DGAFFHVRCACHILNLVAKDGLAVIAGTIEKIKAIVLAVKSSPLQWEELMKCASECDL
                     DKSKGISYDVSTRWNSTYLMLRDALYYKPALIRLKTSDPRRYDAICPKAEEWKMALTL
                     FKCLKKFFDLTELLSGTQYSTANLFYKGFCEIKDLIDQWCVHEKFVIRRMAVAMSEKF
                     EKYWKVSNIALAVACFLDPRYKKILIEFYMKKFHGDSYKVHVDDFVRVIRKLYQFYSS
                     CSPSAPKTKTTTNDSMDDTLMENEDDEFQNYLHELKDYDQVESNELDKYMSEPLLKHS
                     GQFDILSWWRGRVAEYPILTQIARDVLAIQVSTVASESAFSAGGRVVDPYRNRLGSEI
                     VEALICTKDWVAASRKGATYFPTMIGDLEVLDSVIAAATNHENHMDEDEDAIEFSKNN
                     EDVASGSSP"
     intron          1112. .1218
                     /number=1
     CDS             1363. .1671
                     /codon_start=1
                     /product="ORFb"
                     /protein_id="CAA29006.1"
                     /db_xref="GI:22114"
                     /db_xref="GOA:P08771"
                     /db_xref="UniProt/Swiss-Prot:P08771"
                     /translation="MQMVQLQIRVKMIWLLFMNHNHNHNHNQNHNHSHNLNPKKKHHR
                     RGQRSAHRMYGSISPRRKLKWRSMERNTFRYGDIATFLIARLSIGLRVIMEQADFEIT
                     "
     misc_feature    1589. .4246
                     /note="pcAcXY"
     intron          2567. .2637
                     /number=2
     misc_feature    2754. .4302
                     /note="pcAcQ"
     intron          3435. .3523
                     /number=3
     intron          3616. .4002
                     /number=4
     polyA_site      4302
ORIGIN

Alignment Scores:
Pred. No.:              1.01e+03        Length:         4565
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x ZMAC1 (1-4565)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1420 CACAACCACAACCACAAC 1437

RESULT 307
ZMTAC/c
LOCUS       ZMTAC                   4565 bp    DNA     linear   PLN 07-JUL-2002
DEFINITION  Zea mays transposable element Ac.
ACCESSION   X01380
VERSION     X01380.1  GI:22489
KEYWORDS    transposon; unidentified reading frame.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 4565)
  AUTHORS   Mueller-Neumann,M., Yoder,J.I. and Starlinger,P.
  TITLE     The DNA sequence of the transposable element Ac of Zea mays L
  JOURNAL   Mol. Gen. Genet. 198, 19-24 (1984)
REFERENCE   2
  AUTHORS   Kunze,R. and Starlinger,P.
  TITLE     The putative transposase of transposable element Ac from Zea mays
            L. interacts with subterminal sequences of Ac
  JOURNAL   EMBO J. 8 (11), 3177-3185 (1989)
   PUBMED   2555157
REFERENCE   3  (bases 1 to 4565)
  AUTHORS   Kunze,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAR-1986)
COMMENT     Data kindly reviewed (20-AUG-1985) by M. Mueller-Neumann.
FEATURES             Location/Qualifiers
     source          1. .4565
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4577"
     repeat_region   1. .4565
                     /transposon="Ac element"
     repeat_unit     1. .11
                     /note="imperfect inverted repeat A"
     polyA_signal    complement(992. .996)
                     /note="ORF1"
     polyA_signal    complement(1021. .1026)
                     /note="ORF1"
     CDS             complement(1064. .1729)
                     /note="unnamed protein product; ORF1"
                     /codon_start=1
                     /protein_id="CAA25635.1"
                     /db_xref="GI:22490"
                     /db_xref="UniProt/TrEMBL:Q41857"
                     /translation="MAVAMSEKFEKYWKVSNIALAVACFLDPRYKKILIEFYMKKFHG
                     DSYKVHVDDFVRVIRKLYQFYSSCSPSAPKTKTTTNDSMDDTLMENEDDEFQNYLHEL
                     KDYDQVESNELDKYMSEPLLKHSGQFDILSWWRGRVAEYPILTQIARDVLAIQVSTVA
                     SESAFSAGGRVVDPYRNRLGSEIVEALICTKDWVAASRKGECIYVIMKFQFIVIQQLF
                     YLY"
     polyA_signal    complement(1750. .1753)
                     /note="ORF2"
     polyA_signal    complement(1761. .1765)
                     /note="ORF2"
     polyA_signal    complement(1779. .1784)
                     /note="ORF2"
     TATA_signal     complement(1800. .1806)
     CDS             complement(1882. .3165)
                     /note="unnamed protein product; ORF2"
                     /codon_start=1
                     /protein_id="CAA25636.1"
                     /db_xref="GI:22491"
                     /db_xref="GOA:Q9SAP9"
                     /db_xref="UniProt/TrEMBL:Q9SAP9"
                     /translation="MAIVHEPQPQPQPQPEPQPQPQPEPEEEAPQKRAKKCTSDVWQH
                     FTKKEIEVEVDGKKYVQVWGHCNFPNCKAKYRAEGHHGTSGFRNHLRTSHSLVKGQLC
                     LKSEKDHGKDINLIEPYKYDEVVSLKKLHLAIIMHEYPFNIVEHEYFVEFVKSLRPHF
                     PIKSRVTARKYIMDLYLEEKEKLYGKLKDVQSRFSTTMDMWTSCQNKSYMCVTIHWID
                     DDWCLQKRIVGFFHVEGRHTGQRLSQTFTAIMVKWNIEKKLFALSLDNASANEVAVHD
                     IIEDLQDTDSNLVCDGAFFHVRCACHILNLVAKDGLAVIAGTIEKIKAIVLAVKSSPL
                     QWEELMKCASECDLDKSKGISYDVSTRWNSTYLMLRDALYYKPALIRLKTSDPRRYVC
                     LNCCTCHHYKFSINQMSIIVGTMQFVLKPRSGRWH"
     TATA_signal     complement(3221. .3228)
                     /note="ORF3
                     putative"
     TATA_signal     3557. .3565
                     /note="ORF3
                     putative"
     CDS             3739. .4203
                     /note="unnamed protein product; ORF3"
                     /codon_start=1
                     /protein_id="CAA25637.1"
                     /db_xref="GI:1666226"
                     /db_xref="UniProt/TrEMBL:Q41858"
                     /translation="MTQQLGAMVLTCCTAKCCEPVSSRGDRETADGRMGERRAVEVWR
                     TADRRWRMADGRTADGRAVEWRSGRMGGPRRGWASREQWSGGRTVDGGVCVRRAASRP
                     SPRVVPGAAQRPAGWETGRVGESRRERDASPASACGWRRPDSGVGAWRRVNV"
     polyA_signal    4496. .4501
                     /note="ORF3
                     putative"
     polyA_signal    4522. .4526
                     /note="ORF3
                     putative"
     repeat_unit     4555. .4565
                     /note="imperfect direct repeat A'"
```

```
ORIGIN

Alignment Scores:
Pred. No.:              1.01e+03        Length:         4565
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x ZMTAC (1-4565)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       3146 CACAACCACAACCACAAC 3129

RESULT 308
YLACO5/c
LOCUS       YLACO5                  4570 bp    DNA     linear   PLN 15-APR-2005
DEFINITION  Yarrowia lipolytica ACO5 gene.
ACCESSION   AJ001303
VERSION     AJ001303.1  GI:3184145
KEYWORDS    ACO5 gene; Acyl-CoA oxidase 5.
SOURCE      Yarrowia lipolytica
  ORGANISM  Yarrowia lipolytica
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Dipodascaceae; Yarrowia.
REFERENCE   1
  AUTHORS   Le Clainche,A.
  JOURNAL   Thesis (1997) Institut National Agronomique Paris-Grignon
REFERENCE   2  (bases 1 to 4570)
  AUTHORS   Nicaud,J.M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-AUG-1997) Nicaud J.M.J., Laboratoire de Genetique
            Moleculaire et Cellulaire, Institut National Agronomique
            Paris-Grignon, Thiverval-Grignon, 78850, FRANCE
FEATURES             Location/Qualifiers
     source          1. .4570
                     /organism="Yarrowia lipolytica"
                     /mol_type="genomic DNA"
                     /strain="W29"
                     /db_xref="taxon:4952"
     gene            1805. .3904
                     /gene="ACO5"
     CDS             1805. .3904
                     /gene="ACO5"
                     /EC_number="1.3.3.6"
                     /codon_start=1
                     /product="Acyl-CoA oxidase 5"
                     /protein_id="CAA04663.1"
                     /db_xref="GI:3184146"
                     /db_xref="GOA:O74938"
                     /db_xref="InterPro:IPR001005"
                     /db_xref="InterPro:IPR002655"
                     /db_xref="InterPro:IPR006090"
                     /db_xref="InterPro:IPR006091"
                     /db_xref="UniProt/TrEMBL:O74938"
                     /translation="MNNNPTNVILGGKEYDTFTEPPAQMELERAKTQFKVRDVTNFLT
                     GSEQETLLTERIMREIERDPVLNVAGDYDADLPTKRRQAVERIGALARYLPKDSEKEA
                     ILRGQLHGIVDMGTRTRIAVHYGLFMGAIRGSGTKEQYDYWVAKGAATLHKFYGCFAM
                     TELGHGSNVAGLETTATLDKDTDEFIINTPNSGATKWWIGGAAHSATHTACLARLIVD
                     GKDYGVKIFIVQLRDLNSHSLLNGIAIGDIGKKMGRDAIDNGWIQFTDVRIPRQNMLM
                     RYDRVSRDGEVTTSELAQLTYGALLSGRVTMIAESHLLSARFLTIALRYACIRRQFGA
                     VPDKPETKLIDYPYHQRRLLPLLAYTYAMKMGADEAQQQYNSSFGALLKLNPVKDAEK
                     FAVATADLKALFASSAGMKAFTTWAAAKIIDECRQACGGHGYSGYNGFGQAYADWVVQ
                     CTWEGDNNVLCLSMGRSLIQSCIAMRKKKGHVGKSVEYLQRRDELQNARVDNKPLTDP
                     AVLITAWEKVACEAINRATDSFIKLTQEGLSPDQAFEELSQQRFECARIHTRKHLITS
                     FYARISKAKARVKPHLTVLANLFAVWSIEEDSGLFLREGCFEPAEMDEITALVDELCC
                     EAREQVIGFTDAFNLSDFFINAPIGRFDGDAYKHYMDEVKAANNPRNTHAPYYETKLR
                     PFLFRPDEDEEICDLDE"
ORIGIN

Alignment Scores:
Pred. No.:              1.01e+03        Length:         4570
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x YLACO5 (1-4570)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        916 CACAATCACAATCACAAT 899

RESULT 309
AF082296
LOCUS       AF082296                4600 bp    DNA     linear   PLN 25-APR-2001
DEFINITION  Ustilago maydis DNA binding protein Ncp1 gene, complete cds.
ACCESSION   AF082296
VERSION     AF082296.1  GI:4234915
KEYWORDS    .
SOURCE      Ustilago maydis
  ORGANISM  Ustilago maydis
            Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
            Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
REFERENCE   1  (bases 1 to 4600)
  AUTHORS   Hartmann,H.A., Kruger,J., Lottspeich,F. and Kahmann,R.
  TITLE     Environmental signals controlling sexual development of the corn
            Smut fungus Ustilago maydis through the transcriptional regulator
            Prf1
  JOURNAL   Plant Cell 11 (7), 1293-1306 (1999)
   PUBMED   10402430
REFERENCE   2  (bases 1 to 4600)
  AUTHORS   Hartmann,H.A., Kruger,J., Lottspeich,F. and Kahmann,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-AUG-1998) AG Kahmann, Inst. f. Genetik, LMU,
            Maria-Ward Str. 1a, Muenchen 80638, Germany
FEATURES             Location/Qualifiers
     source          1. .4600
                     /organism="Ustilago maydis"
                     /mol_type="genomic DNA"
                     /strain="FBD11"
                     /db_xref="taxon:5270"
     gene            2525. .3727
                     /gene="ncp1"
     CDS             2525. .3727
                     /gene="ncp1"
                     /function="nutritional control of prf1"
                     /note="binds to the prf1 gene promoter"
                     /codon_start=1
                     /product="DNA binding protein Ncp1"
                     /protein_id="AAD13020.1"
                     /db_xref="GI:4234916"
                     /translation="MTDLARNGSILVAEIKAIHPANGSDATNGAVPHANGDANGSAIL
                     SKDAPTLPATAGAGAAAPIESSITDPATQPHLGNGHAHQATDLSIGAGAVAVPHQADT
                     SVASGIADPVAQSPAPTATDAAAPGLVQPQTTAIQPVQPNDHAREVSHINTAVPAAGV
                     GAAAATAAPGSAVEPTSAVSAVAPSTAPGSAISAQQPQQLNAVASSPTAAGAPSFPPG
                     TIVPSSVTGAGAGAGAGAGAGAGAATVAPIVAHPPPTTLAATEAPKSALSPKEAKKMD
                     KLMKKEAKNEHKSLKRAIKDAAASEKLFRKSRDAETQALKRHQKAQTKEHKAAKALSK
                     VKAEYEKYAADLAKAVEDLEIKKRHTESTREGYETSKKHIDELRNQKTVNDQARSRQA
                     AELHSARR"
ORIGIN

Alignment Scores:
Pred. No.:              1.02e+03        Length:         4600
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x AF082296 (1-4600)

Qy          1 HisAsnHisAsnHisAsn 6
```

```
                    ||||||||||||||||||
Db       1237 CACAATCACAATCACAAT 1254

RESULT 310
AF316534
LOCUS       AF316534                4633 bp    DNA     linear   PLN 01-AUG-2002
DEFINITION  Prunus persica ethylene-responsive sensor (ERS) gene, complete cds.
ACCESSION   AF316534
VERSION     AF316534.1  GI:22036193
KEYWORDS    .
SOURCE      Prunus persica (peach)
  ORGANISM  Prunus persica
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Rosales; Rosaceae; Amygdaloideae; Prunus.
REFERENCE   1  (bases 1 to 4633)
  AUTHORS   Bonghi,C., Rasori,A., Tonutti,P. and Ramina,A.
  TITLE     Ethylene receptors in peach
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4633)
  AUTHORS   Bonghi,C., Rasori,A., Tonutti,P. and Ramina,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-2000) Enviromental Agronomy and Crop Science,
            University of Padova, via Romea, 16 Agripolis, Legnaro, Padova
            35020, Italy
FEATURES             Location/Qualifiers
     source          1. .4633
                     /organism="Prunus persica"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3760"
                     /note="authority: Prunus persica L. Batch"
     gene            <1652. .>4255
                     /gene="ERS"
                     /note="ERS-like ethylene receptor"
     mRNA            join(<1652. .2557,2770. .3138,3226. .3492,3672. .3799,
                     3991. .>4255)
                     /gene="ERS"
                     /product="ethylene-responsive sensor"
     CDS             join(1652. .2557,2770. .3138,3226. .3492,3672. .3799,
                     3991. .4255)
                     /gene="ERS"
                     /codon_start=1
                     /product="ethylene-responsive sensor"
                     /protein_id="AAM89517.1"
                     /db_xref="GI:22036194"
                     /translation="MDSCDCIDAQWPPDELLVKYQYISDVFIALAYFSIPLELIYFVQ
                     KSSFFPYRWVLMQFGAFIVLCGATHFINVWTFYMHSKTVAVVMTVAKVSCAVVSCATA
                     LMLVHIIPDLLSVKTRELFLKNRAEELDREMGRMLTQEETGRHVRMLTHEIRSTLDRH
                     TILKTTLVELGRTLGLQECALWMPSRTGMNLQLSHTLNYQIQVGSTVPINLPIINEVF
                     TSSRAIRIPYTCPLARIRPLVGRYVPPEVIALRVPLLNLSNFQINDWPDLSAKSYAIM
                     VLILPTDSTRKWRDHELELVDVVADQVAVALSHAAILEESMRARDQLMEQNVALDLAR
                     REAEMAIHARNDFLAVMNHEMRTPMHAIIALSSLLLETELISEQRVMIETVLKSSNLL
                     ETLINDVLDLSRLEDGSLELDIRKFNLHGVFREVINLIQPITSVKKLSMSLILSPDLP
                     LCAVGDENRLMQTILNVAGNAVKFTKEGYISIVASVAKPESSRDWRPPEFYPASTDGH
                     FYLRVQVKDSGCGILPQDIPHLFTKFSQPRSGSNRINDGAGLGLAICRRFINTMGGHI
                     WIESEGIDKGSIVAFIVKLGICDNPCDTTAAHQMAPQALPQPQPQPNYGSGDRIAHRP
                     TVSFRDGGYTSSNPRYQRSI"
ORIGIN

Alignment Scores:
Pred. No.:              1.02e+03        Length:         4633
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x AF316534 (1-4633)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       4137 CACAACCACAACCACAAC 4154

RESULT 311
CQ596381
LOCUS       CQ596381                4779 bp    DNA     linear   PAT 02-FEB-2004
DEFINITION  Sequence 24139 from Patent WO0171042.
ACCESSION   CQ596381
VERSION     CQ596381.1  GI:41652900
KEYWORDS    .
SOURCE      Drosophila sp.
  ORGANISM  Drosophila sp.
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1
  AUTHORS   Venter,J.C., Adams,M., Li,P.W. and Myers,E.W.
  TITLE     Detection kits, such as nucleic acid arrays, for detecting the
            expression of 10,000 or more Drosophila genes and uses thereof
  JOURNAL   Patent: WO 0171042-A 24139 27-SEP-2001;
            PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1. .4779
                     /organism="Drosophila sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:7242"
ORIGIN

Alignment Scores:
Pred. No.:              1.06e+03        Length:         4779
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x CQ596381 (1-4779)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       2493 CACAACCATAACCACAAC 2510

RESULT 312
MZEAC9WX/c
LOCUS       MZEAC9WX                4810 bp    DNA     linear   PLN 28-FEB-1994
DEFINITION  Maize transposable element Activator (Ac9) from the waxy locus.
ACCESSION   K01964
VERSION     K01964.1  GI:168402
KEYWORDS    activator; repeat region; transposase; transposon.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 4810)
  AUTHORS   Pohlman,R.F., Fedoroff,N.V. and Messing,J.
  TITLE     The nucleotide sequence of the maize controlling element Activator
  JOURNAL   Cell 37 (2), 635-643 (1984)
   PUBMED   6327080
REFERENCE   2  (bases 1 to 4810)
  AUTHORS   Pohlman,R.F., Fedoroff,N.V. and Messing,J.
  TITLE     Correction: nucleotide sequence of Ac
  JOURNAL   Cell 39 (2 Pt 1), 417 (1984)
   PUBMED   6094018
COMMENT     Original source text: Maize DNA, clone Ac wx-m9.
            [2]  revises [1].
            The ends of the Ac9 element comprise an 11 bp imperfect terminal
            repetition and the insertion site within the wx locus is flanked by
            an 8 bp directly repeated sequence.
FEATURES             Location/Qualifiers
     source          1. .4810
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4577"
     misc_feature    114. .121
                     /note="insertion target sequence copy A"
```

```
     misc_feature    122. .4682
                     /standard_name="transposable element Ac9"
     repeat_region   122. .132
                     /note="inverted repeat copy A (imperfect)"
     repeat_region   4672. .4682
                     /note="inverted repeat copy B (imperfect)"
     misc_feature    4683. .4690
                     /note="insertion target sequence copy B"
ORIGIN      5 bp upstream of PstI site.

Alignment Scores:
Pred. No.:              1.06e+03        Length:         4810
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x MZEAC9WX (1-4810)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       3267 CACAACCACAACCACAAC 3250

RESULT 313
AF275723
LOCUS       AF275723               4911 bp    mRNA    linear   INV 14-DEC-2000
DEFINITION  Dictyostelium discoideum Ras guanine nucleotide exchange factor
            RasGEFB (gefB) mRNA, complete cds.
ACCESSION   AF275723
VERSION     AF275723.1  GI:11078660
KEYWORDS    .
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 4911)
  AUTHORS   Wilkins,A., Chubb,J.R. and Insall,R.H.
  TITLE     A novel Dictyostelium RasGEF is required for normal endocytosis,
            cell motility and multicellular development
  JOURNAL   Curr. Biol. 10 (22), 1427-1437 (2000)
   PUBMED   11102804
REFERENCE   2  (bases 1 to 4911)
  AUTHORS   Wilkins,A. and Insall,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUN-2000) Biosciences, Birmingham University,
            Edgbaston, Birmingham B15 2TT, UK
FEATURES             Location/Qualifiers
     source          1. .4911
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX3"
                     /db_xref="taxon:44689"
                     /clone="SSK260 from Tsukuba cDNA project"
     gene            1. .4911
                     /gene="gefB"
     CDS             296. .4885
                     /gene="gefB"
                     /codon_start=1
                     /product="Ras guanine nucleotide exchange factor RasGEFB"
                     /protein_id="AAG29138.1"
                     /db_xref="GI:11078661"
                     /translation="MVVILWGWNENRFNTLDQVISFDFIEENKKLYLKILENEKQQKI
                     YNISNLIGRRMDNYRDNNNNNGNGSNINNNGIIINSSMGSQYFNLNKGIIIVESDNET
                     FLIEEIDYLGKFPILKQFIKCEEFIGISTLEIISNIEKQLSNIVNLKSNTTEQTDRKY
                     KTNLIDFKESIIQLEKDCKNLLKQSNLINPSIKYSVVKSVKFLLKEIKFEFYHGDINI
                     KRIELFQQLKKFLKIHFWNNPNLILNSGSNIDMTDSNDSISSEEINSPNDLNLSLSSF
                     PSISSPSNSSYNNINTLTFSNNNNNNNNNSNNNNNNNNNNNNNNSIINSNNNNNSNNN
                     SNISNSSNNNNNNSNNSNNISINVSKTGNNSVINSPTQSSSCYIDCDDTMSNYTDCTN
                     TTNTNSTTTTTTTTTTYGNNLNVYNNNNNNNTNNNNNNVPYKKPALKMMKSPSFKIQL
                     IQEVYLEIMDQCKKIQNYLYELLDTYLTDEEYINNNNNNNNKDISILFENNTSYCIAQ
                     EIFISLIKNSNDFLYQVQLIDLTDTRKKSKELSEVSNQLFYSTQQLFPSISDHLDFIE
                     MDIFTNNNTTTTTTTTTTTTNTTTNPSNTIKSSHSSLLSSSYSSSKDDDIVLTINKVL
                     SKLDTFSKEYTKLIGNNLSSNYYVDSGNNNESVSSPKCGLFLSHSLSTDSPMTYVKRN
                     TSSGSSASTSSYTTPVSSSASLSFSSSTQTPSSAASQHHIQQILQQQQQQQQQQQQQQ
                     QQQQPQQIQSQSQQQHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQIQSQSQQQQQLNRK
                     PHYSSISPTSMLVYNATGGKFGVVNQNNNSAESTPTFQLNTMDILNSLMGKEGQNSNT
                     NNANNPNNNNNNNNNNNNNNNNNNNNNNNNNENKNENKNETNKSGHNYNNSSTSTLSSS
                     TTIGDLANVSSTNSPNSSTPNLLLPPHQFHNHNREYNHSHHHLLSSSSNNVMINNRSI
                     GKSFSTGNFSQPQPIMPSNGIGGSGGGGLGTTLSVGNGGIVGGLQSRSSSFEENQLQY
                     AIESFMEVYNHLTSKDKVDTSIWLEEEKEDINVYYINNTNIDGKFYRSIKYASLNQLI
                     LKLTCESNTELRFTKTFITTYRSFTTGDIFLMKLIQRYFTPNLKNIVPYQFVQKIQIP
                     IQLRVLNVLRLWIEQYPSDFQQPPQITTQSTQPIQNSTTQPQPQPQPQQPQPQLQQST
                     NLQNSTIQQPSLFKMLTAFLNSTRRNGHGQYSDLILKKFKQQKQKDRLQLPIINNGQG
                     IPKAKIFWMKYSTEFIFSQSSTDIAEQLTLLDFDSYKSIEEIELLNQAWSKPDQKTNT
                     PNIASMVNRFNSFSSFVSWAILRENDIKQRSKMMLKMIKICYALYKLSNFNGLLAGLS
                     GLMASGVYRLNHTKSLISKQYQKKFDFLCKFLDTKKSHKVYRDIIHSTCPPLIPYLGI
                     YLTDLTFIEDGNQDEIKGLINFKKREMIFNTILEIQQYQQQGYTIKPKPSVLGFLCEL
                     PHMSDKKKFEDDTYEQSILLEPKNSTKLEVMNAKK"
ORIGIN

Alignment Scores:
Pred. No.:              1.08e+03        Length:         4911
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-858-332G-28 (1-6) x AF275723 (1-4911)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       3849 CACAACCACAACCACAAC 3866

RESULT 314
AX594148
LOCUS       AX594148               4938 bp    DNA     linear   PAT 13-FEB-2003
DEFINITION  Sequence 1 from Patent WO02083910.
ACCESSION   AX594148
VERSION     AX594148.1  GI:28375378
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Farmer,A.A.
  TITLE     Sequence specific recombinase-based methods for producing intron
            containing vectors and compositions for use in practicing the same
  JOURNAL   Patent: WO 02083910-A 1 24-OCT-2002;
            Clontech Laboratories Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .4938
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Synthetic Vector"
ORIGIN

Alignment Scores:
Pred. No.:              1.09e+03        Length:         4938
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x AX594148 (1-4938)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        141 CATAATCATAATCATAAT 158

RESULT 315
AF076279
LOCUS       AF076279               5015 bp    DNA     circular INV 12-JUL-1999
DEFINITION  Dictyostelium firmibasis plasmid Dfp1, complete plasmid sequence.
```

```
ACCESSION   AF076279
VERSION     AF076279.1  GI:3386371
KEYWORDS    .
SOURCE      Dictyostelium firmibasis
  ORGANISM  Dictyostelium firmibasis
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 5015)
  AUTHORS   Gonzales,C.M., Spencer,T.D., Pendley,S.S. and Welker,D.L.
  TITLE     Dgp1 and Dfp1 are closely related plasmids in the Dictyostelium
            Ddp2 plasmid family
  JOURNAL   Plasmid 41 (2), 89-96 (1999)
   PUBMED   10087212
REFERENCE   2  (bases 1 to 5015)
  AUTHORS   Welker,D.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUL-1998) Biology, Utah State University, Logan, Utah
            84322-5305, USA
FEATURES             Location/Qualifiers
     source          1. .5015
                     /organism="Dictyostelium firmibasis"
                     /mol_type="genomic DNA"
                     /strain="CR II 2B"
                     /db_xref="taxon:79012"
                     /plasmid="Dfp1"
     gene            1. .2625
                     /gene="rep"
     CDS             1. .2625
                     /gene="rep"
                     /note="similar to plasmid Ddp2 Rep protein"
                     /codon_start=1
                     /product="Rep protein"
                     /protein_id="AAC33156.1"
                     /db_xref="GI:3386372"
                     /translation="MDENNKLLPWYNYVFFFQRILEEYLPSKRIKDTDVMRHYYTPRV
                     LVMKEIFRTIINKHYYPKEGPLRCERDCYDPKVIFPKFSLEFDMEAPFTDLDDLFKMK
                     KTKNNAKIIVSDGDQTIFIIDHFSRIADKQIFCKKTIASFRNAEKIVTSAGYKLEDYR
                     QLGIDWYGFINKLHTACCIEEQGEDMDVFAGIFEYPDLIAILTLVNDIIVEERNGYEE
                     DLSLLYTLYPFTNLEEESNTTTTTTTTTSKRGRGRVPITTPTTTTTTTTTTTTTAFEP
                     LKKRRVMGSRSSGSLKEVRVNNIALSSTPSDTSDNEVDPETSIVSISSRIKDCILETI
                     NSKGELDTRTVKAIFSSLQNKDYSLDLVDKIFQQNKSEKVITVSNKMYSLVAKLDFNE
                     IKISDFSEELFDLSKRISFEKSTNILLATQNADNGFGFLWIPIVDNHTSTSIYLSPIN
                     SKVDHDRIRLILKFVHLCIALGDINRFISSRSLSFERFKQVSKEILGMSQRILSLEAD
                     VNTIKQLFSKKKGKMDKESVEFTNNIIIETELAVYVNEFINNNIYIKLSNLDNALKLS
                     NPPVSSLVVEYDPCNIDHKVGFLYHCRSEISKFDNNKNSYSMNKLLTSFTPSNIKNVS
                     LENEEDIKRKYHLMSSELSEILKDPTPFIPTSFKRYMNVTFTNTAYNYNSVNTFDDIS
                     EGISITNLRNIFAEDSPEENRYLEYIGGTRVIRAFFFAPCLIKITNIAFVSEQLISNR
                     SISRQIKSIKIKNHSYIPIEIKVGNCIVDTVKSDETREVFINSSEFSFSISCLNISFS
                     TTLISKSKLKNIKTIIPEKYQQMTDYDRYTTESNRLMDNFLTILKDTNKLSKLDEYLC
                     KYFDLSIGKEDTDESSDE"
     repeat_region   3053. .3703
                     /note="repeat elements DR1, DR2, DR3 and DR4"
                     /rpt_type=direct
                     /rpt_unit="3053. .3217"
     repeat_region   3719. .4777
                     /note="repeat elements IR1 and IR2"
                     /rpt_type=inverted
                     /rpt_unit="3719. .4193"
ORIGIN

Alignment Scores:
Pred. No.:                1.11e+03        Length:         5015
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       2               Gaps:           0

US-09-858-332G-28 (1-6) x AF076279 (1-5015)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        783 CACAACCACAACCACAAC 800
```

```
RESULT 316
AR097042
LOCUS       AR097042                5163 bp    DNA     linear   PAT 14-FEB-2001
DEFINITION  Sequence 4 from patent US 6071518.
ACCESSION   AR097042
VERSION     AR097042.1  GI:12805772
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5163)
  AUTHORS   Petersen,C.
  TITLE     GP900 glycoprotein and fragments for treatment and
            detection/diagnosis of cryptosporidium
  JOURNAL   Patent: US 6071518-A 4 06-JUN-2000;
FEATURES             Location/Qualifiers
     source          1. .5163
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

Alignment Scores:
Pred. No.:                1.14e+03        Length:         5163
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x AR097042 (1-5163)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        891 CACAACCACAACCACAAC 908

RESULT 317
AR279261
LOCUS       AR279261                5163 bp    DNA     linear   PAT 10-APR-2003
DEFINITION  Sequence 4 from patent US 6514697.
ACCESSION   AR279261
VERSION     AR279261.1  GI:29713989
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5163)
  AUTHORS   Petersen,C., Barnes,D.A., Nelson,R.C. and Gut,J.
  TITLE     Methods for detection of Crytosporidium species and isolates and
            for diagnosis of Cryptosporidium infections
  JOURNAL   Patent: US 6514697-A 4 04-FEB-2003;
            The Regents of the University of California; Oakland, CA
FEATURES             Location/Qualifiers
     source          1. .5163
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:                1.14e+03        Length:         5163
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x AR279261 (1-5163)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        891 CACAACCACAACCACAAC 908

RESULT 318
AF253047
```

```
LOCUS       AF253047               5230 bp    DNA     linear   PLN 13-NOV-2002
DEFINITION  Candida albicans DNA primase large subunit Pri2p (pri2) gene,
            partial cds; large subunit Cp complex Cdc68p (cdc68) gene, complete
            cds; and Cso1p (cso99) gene, partial cds.
ACCESSION   AF253047
VERSION     AF253047.1  GI:12082668
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 5230)
  AUTHORS   Buurman,E.T., Jiang,W., McCoy,M., Averett,D.R., Thompson,C.M. and
            Wobbe,C.R.
  TITLE     Validation of Cdc68p as a novel antifungal target
  JOURNAL   Arch. Microbiol. 178 (6), 428-436 (2002)
   PUBMED   12420162
REFERENCE   2  (bases 1 to 5230)
  AUTHORS   Buurman,E.T., Jiang,W. and Wobbe,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-APR-2000) Scriptgen Pharmaceuticals, 610 Lincoln
            Street, Waltham, MA 02451, USA
FEATURES             Location/Qualifiers
     source          1. .5230
                     /organism="Candida albicans"
                     /mol_type="genomic DNA"
                     /strain="1006"
                     /db_xref="taxon:5476"
     gene            complement(<1. .>379)
                     /gene="pri2"
     mRNA            complement(<1. .>379)
                     /gene="pri2"
                     /product="DNA primase large subunit Pri2p"
     CDS             complement(<1. .379)
                     /gene="pri2"
                     /codon_start=1
                     /transl_table=12
                     /product="DNA primase large subunit Pri2p"
                     /protein_id="AAG48573.1"
                     /db_xref="GI:12082669"
                     /translation="MFRQVKRRTAGRRNFEDSTSTATTTTTTYTNGEKDSPKYLTPS
                     ALYASRLSFYDLPPTQEITLEEFETWAIDRLKILIEIESCLARSKTLKEIETSIKPLL
                     LKYLPLNPTSGGGTVNYEKERMKD"
     gene            <834. .>3992
                     /gene="cdc68"
     mRNA            <834. .>3992
                     /gene="cdc68"
                     /product="large subunit Cp complex Cdc68p"
     CDS             834. .3992
                     /gene="cdc68"
                     /note="Spt16p"
                     /codon_start=1
                     /transl_table=12
                     /product="large subunit Cp complex Cdc68p"
                     /protein_id="AAG48574.1"
                     /db_xref="GI:12082670"
                     /translation="MSEVNIDAGLFYKRLSIFQKQLTANNIPQALIIVGARSDDNTYK
                     KSTVLQNWLLGYEFIHTAIYITDKKCIFITSEGKSKHLKHLTNQKPDLVELWIRTKDV
                     EHNKQLFIKLLETMTKLDSKYGKILKDKYDGKFIDEWNQILNDNNNNTTNDHALSAVD
                     LAVTVSQALAVKDSEEFNNTKIASNASVVMMDTFVNDMMIIVDDEKKITNSQLTDQIE
                     DKIENNKWYLKTKLGKNLLQSIKDFDPEYLEYCYSPIIQSGGDYDLKPSAVSTDKPLI
                     GEGVILSSIGLRYKSYCSNIARTFLIDPTSEMETNYDFLLQLQKYIVDNLLKDGVPAS
                     KVYQDTIDYIKKERPDLVNHFTKNCGWLLGMEFRDSTFILNAKTTDRKLTTGQIISLT
                     IGFNNLSNDKNDKNDNKTNHQKNKQTYALLLTDTIKITDDSSILLTNYSKDRAAISFS
                     FNDDNETQKENNNNNNKRPGLSQTSNTTGLKLESTENTAILKSKLRHENTNADDANSE
                     KLRQEIQIKLHEKRLQEGLARFSKADATDADDFKPIFKKYESYVRESQIPNSVNDLKI
                     HIDYKNQTIILPISGRPVPFHINSYKSGSQNEEGDFTYLRLNFNSPGAGGNVTKKQEL
                     PYEDSPDNSFLRSITIRSRDRQRMVDVYKAIQDLKKDSVKREQEKKQMADVITQANLI
                     ELKGSRVKKLNNVFIRPTPDTKKIGGVLQIHENGLRYQSQPQSQSQSNFKNDQRVDVL
                     FSNIKHLFFQPCKDELIVLIHCHLKNPIMIGKRKTFDVQFYREASDMAFDETGGRKRK
                     YRYGDEDELQQEQEERRRKALLDKEFKGFAELIADSSHGMVDLDIPFRELGFQGVPFR
                     SSVLCVPTRDCLVQLIDPPYLVVTLEEIEIAHLERVQFGLKNFDLVFVFKDFNKPVVH
                     INTIPVELLEDVKSWLTDVDIPISEGQMNLNWVQIMKTVLADPYQFFIDGGWAFLTGQ
                     GESDEEEESDEESDFRVSDEDPQDEDEESDDYASEEESDDYSGSDDDGSGGGDDDDSE
                     SGEDWDALERKAAKADRNSGFD"
     gene            complement(<4198. .>5230)
                     /gene="cso99"
     mRNA            complement(<4198. .>5230)
                     /gene="cso99"
                     /product="Cso1p"
     CDS             complement(4198. .>5230)
                     /gene="cso99"
                     /codon_start=2
                     /transl_table=12
                     /product="Cso1p"
                     /protein_id="AAG48575.1"
                     /db_xref="GI:12082671"
                     /translation="IKYFNITTTTQPIMSSISGISNISSISKLNGNSNKPTTLLPISI
                     LLNLQKILKKSIFSRKFYQEINDKVLSKSSTVDQNLYFICYFSLLISAILNNKPKILY
                     FLKNQWYKFIQLVKSATGISTTTTATATTTATKEKEENNGGSTSYLEKILNPKPPVYT
                     EGTPSKLAYHLKKINSYLADIRIFNRLTDSIKYMPWIIDEYHAFTNTNSPTSIPKFDR
                     FINLLQSLNCLVLEILENLGWLTDHDWIGTGNNEWWCIETYIWCCRVWGVYLLIEIME
                     IIRRTPMTKWSNNKQWQINLFKNVIQLPLVLHWSLRDGCLTPFWVGLCGSGASWWNFK
                     DMWSSIDLS"
ORIGIN

Alignment Scores:
Pred. No.:                1.15e+03        Length:         5230
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       15              Gaps:           0

US-09-858-332G-28 (1-6) x AF253047 (1-5230)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       2995 CACAACCACAATCACAAT 3012

RESULT 319
CQ576536
LOCUS       CQ576536               5239 bp    DNA     linear   PAT 02-FEB-2004
DEFINITION  Sequence 4294 from Patent WO0171042.
ACCESSION   CQ576536
VERSION     CQ576536.1  GI:41639670
KEYWORDS    .
SOURCE      Drosophila sp.
  ORGANISM  Drosophila sp.
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1
  AUTHORS   Venter,J.C., Adams,M., Li,P.W. and Myers,E.W.
  TITLE     Detection kits, such as nucleic acid arrays, for detecting the
            expression of 10,000 or more Drosophila genes and uses thereof
  JOURNAL   Patent: WO 0171042-A 4294 27-SEP-2001;
            PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1. .5239
                     /organism="Drosophila sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:7242"
ORIGIN

Alignment Scores:
Pred. No.:                1.15e+03        Length:         5239
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x CQ576536 (1-5239)

Qy          1 HisAsnHisAsnHisAsn 6
```

```
                  ||||||||||||||||||
Db        671 CACAATCACAACCACAAC 688

RESULT 320
DDDOKAGEN
LOCUS       DDDOKAGEN               5284 bp    DNA     linear   INV 18-APR-2005
DEFINITION  D.discoideum dokA gene.
ACCESSION   X96869
VERSION     X96869.1  GI:1237201
KEYWORDS    dokA gene; histidine kinase.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1
  AUTHORS   Schuster,S.C., Noegel,A.A., Oehme,F., Gerisch,G. and Simon,M.I.
  TITLE     The hybrid histidine kinase DokA is part of the osmotic response
            system of Dictyostelium
  JOURNAL   EMBO J. 15 (15), 3880-3889 (1996)
   PUBMED   8670893
REFERENCE   2  (bases 1 to 5284)
  AUTHORS   Schuster,S.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAR-1996) S.C. Schuster, Max-Planck-Institut fuer
            Biochemie Membranbiochemie, Am Klopferspitz 18a, Martinsried,
            82152, FRG
FEATURES             Location/Qualifiers
     source          1. .5284
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX2"
                     /sub_strain="214"
                     /db_xref="taxon:44689"
     gene            44. .5056
                     /gene="dokA"
     CDS             44. .5056
                     /gene="dokA"
                     /codon_start=1
                     /product="histidine kinase"
                     /protein_id="CAA65612.1"
                     /db_xref="GI:1237202"
                     /db_xref="GOA:Q23901"
                     /db_xref="InterPro:IPR000014"
                     /db_xref="InterPro:IPR001128"
                     /db_xref="InterPro:IPR001789"
                     /db_xref="InterPro:IPR003594"
                     /db_xref="InterPro:IPR003661"
                     /db_xref="InterPro:IPR004358"
                     /db_xref="InterPro:IPR005467"
                     /db_xref="UniProt/TrEMBL:Q23901"
                     /translation="MSSPHIELHSQRTLSPQPSSNNFELTGNKSCALSASLNGSIDDL
                     NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNDGDKNDKKLLIITNAPTPP
                     LTPPQPTKTQQTLLELNQKFQEQQQQQQQQQQLQKQNKQQQSQKQQEEPPSSQQEEPP
                     LSQQQQEQEQEQEQEQEQEQEESQSKQKIEGKGGGGEEEECEGGGGGEGEGEEQFKEGDEE
                     ETRVLKLIQSTNNWIYWYRRSSLQINEASPKLIQTPSVDRTLCILDSLPCPIFQIGII
                     ENTENGNITLTGFANERLTTLLSCSSNLLVDSLKSIPSPHYNTVPSLESCNSIPINNL
                     APGLLPNTDPTKILDPYISLYNKKQLKERYSNNNNNTNTNNNNNNNNNNNNNNNNNNN
                     NNNNDTTTTTTTTTTTTTTTTTTTTTTTTTTNESDNNNNNNNNNNNNNNNNNNNNNEE
                     IKQNTKEGEEILHESTNSTGTGNSISRPNSILQCILDSFHEQKTITEKVLLSNFLIRQ
                     QYKSRATIKPFPDGCICIFEYIENINLNYQQPPTSLNDRINSTNKLTNEMELLSLSPD
                     SNHQHVQKLHNHNHNHNHNHNHNTTQRASSTDSPPIHSVSANSLSSMSSNSSVTSACA
                     SLSSNSSSNSNNNSNNASSSNCSSNANSNNTNSSSSNCCCCCNNNNNNNTSSTSSTIS
                     SASSTKSKSYLHQQHKNSKVYSTLLNSFENLSFLLKSTPIIIWRADHNGEMVFFKKSD
                     EIPIDEKKVVGNNFQDFLQWVPQTYRTNFQEMFQNSLKCGKIFEFLFWFQTPQNYVQL
                     FKIAGYPIFHIDGKDCIASKRYLIGWLGVTYNYLINDGAEISIKGSANLDSMYEKFKI
                     IYEMPNIERTKLTNNGISSGGDITNANNNGGSSESSNKILEKKKHLISEAEKECFLKC
                     KETYNILFKLSLLGVMFSTFKGIILDANDMLLQTIGYTRGDLENGKIDWMLLTPPEHF
                     EISARALQELKAKRWCQPIEKAYIHKSGKRVPVLITSAMIDGSTEQCITFVFDLSRYR
                     QAEMAAIEATRLKTQFITNISHELRTPCHGIVGMSQLLLDSQLTNTQRDNIDSIKRST
                     DSLISLINDILDFSKLEYGKVTLENESFELLPMIEEVLDSQATAANRKGIDLIFVMGR
                     DYPVPPVIFGDRNSLKKVLLNLVGNAVKFTETGFVLLEISTDYESGDQISLRFTVKDS
                     GIGIPENKIEQIFVPFGQIDGSFSRKYGGSGLGLSFCKELVALMGGYIRVESGDQEGG
                     KGTTFWFAIKVSISSPSYLPNSVPAANQFFYPEYRPSHYNNVIGNGDSPKNVLIIESN
                     QMVIMSIQSILLSMKCECISASKAITALDLLESSKSTENQIDIIVCSDKSTFVQQILD
                     YVTTEKVILYGVDPNSKYNENSKVYSYLVTPITHSKLISSILLSKNLKSKNSFLTTTN
                     NTNNTNNTNNIENKSSIDSPLSITSTSSSIITPILTSNNLDLNNNNNNNNNSILVSNN
                     GGVDGGNNVPSTLTTIQQSQPKKYILVAEDNDINIKVVVRQLEKLGYTAIVGINGLKA
                     LEIIGSFPICLILLDCQMPQMDGFTCSTILRQIEPTGQRIPIIAMTANDSKDRCFEVG
                     MDDYLSKPVRVDRLQKTLSDWIKTDENGNPTNSYNFYPLSYSLVYNNFIDTQLKKEKN
                     DD"
ORIGIN

Alignment Scores:
Pred. No.:                1.16e+03        Length:          5284
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       2               Gaps:            0

US-09-858-332G-28 (1-6) x DDDOKAGEN (1-5284)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1769 CATAATCATAATCATAAT 1786

RESULT 321
AR097041
LOCUS       AR097041                5318 bp    DNA     linear   PAT 14-FEB-2001
DEFINITION  Sequence 3 from patent US 6071518.
ACCESSION   AR097041
VERSION     AR097041.1  GI:12805771
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5318)
  AUTHORS   Petersen,C.
  TITLE     GP900 glycoprotein and fragments for treatment and
            detection/diagnosis of cryptosporidium
  JOURNAL   Patent: US 6071518-A 3 06-JUN-2000;
FEATURES             Location/Qualifiers
     source          1. .5318
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

Alignment Scores:
Pred. No.:                1.17e+03        Length:          5318
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       6               Gaps:            0

US-09-858-332G-28 (1-6) x AR097041 (1-5318)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        891 CACAACCACAACCACAAC 908

RESULT 322
AR279260
LOCUS       AR279260                5318 bp    DNA     linear   PAT 10-APR-2003
DEFINITION  Sequence 3 from patent US 6514697.
ACCESSION   AR279260
VERSION     AR279260.1  GI:29713988
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5318)
  AUTHORS   Petersen,C., Barnes,D.A., Nelson,R.C. and Gut,J.
  TITLE     Methods for detection of Crytosporidium species and isolates and
            for diagnosis of Cryptosporidium infections
  JOURNAL   Patent: US 6514697-A 3 04-FEB-2003;
            The Regents of the University of California; Oakland, CA
```

```
FEATURES             Location/Qualifiers
     source          1. .5318
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              1.17e+03        Length:         5318
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x AR279260 (1-5318)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        891 CACAACCACAACCACAAC 908

RESULT 323
SCYGL176C
LOCUS       SCYGL176C               5378 bp    DNA     linear   PLN 11-AUG-1997
DEFINITION  S.cerevisiae chromosome VII reading frame ORF YGL176c.
ACCESSION   Z72698 Y13135
VERSION     Z72698.1  GI:1322783
KEYWORDS    .
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 5378)
  AUTHORS   Bruschi,C.V., Coglievina,M., Bertani,I., Klima,R., Zaccaria,P. and
            Delneri,D.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 5378)
  AUTHORS   MIPS.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAY-1996) Data collected by MIPS on behalf of the
            European yeast chromosome VII sequencing project. MIPS at the
            Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
            Martinsried, FRG; E-mail: Mewes@mips.embnet.org
FEATURES             Location/Qualifiers
     source          1. .5378
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4932"
                     /chromosome="VII"
     gene            1. .2665
                     /gene="MPT5"
     intron          <1. .88
                     /gene="MPT5"
                     /note="ORF YGL178w"
                     /number=1
     exon            89. .2665
                     /gene="MPT5"
                     /note="ORF YGL178w"
                     /number=2
     CDS             complement(3509. .5173)
                     /note="unnamed protein product; ORF YGL176c"
                     /codon_start=1
                     /protein_id="CAA96888.1"
                     /db_xref="GI:1322784"
                     /db_xref="SGD:S0003144"
                     /db_xref="UniProt/Swiss-Prot:P46945"
                     /translation="MTPNLPGFYYDRERRRYFRISENQSISTTGTTNQYRKDNIKRQC
                     VEENYDKKFSMIKKKRQQTLQKYKLSLLNPLERAFRPLSYEKYMIGLNMQYASHSLTE
                     GHHSHSSANVKSLNFPHRIQIGVLANCILLVTQEGCFHSKLVFATNKGYVAGFSSLDN
                     FSEENFFTGFSMAELNPMLKYKSEPTDVFKTMKLERTVAIKEGPSHYFYHNVNTRSNV
                     HTFAIFLQDFSSLKLLKIRQVKLKENCQVHDSLVVGDTLIITVNYRCHFYDLIPETFP
                     NPYIFSPAKSSRKHKSRSDITSLSFCLQEDALSPLKKTNTGVFYLGYRNGDSMAIVFT
                     NITNMTLQYSKTNGMTSESRNQPIRNSLKSVVSIKALNNKGLILISGMADKENVQQLV
                     IADTFLEDILTEIPVVSFKTKFLNVTKDTEILEISDDGRYFIYGSTSARDGKGDFEVF
                     CTTLSGNLDYEKSEGGNITLYPIGGMKNYCRLENFQFESIHLHSAFIPPRYVNPFDAV
                     EPLGEESSTSPYDIPEEALSQKICILIRREDDPYNGANIFITSALT"
ORIGIN

Alignment Scores:
Pred. No.:              1.18e+03        Length:         5378
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x SCYGL176C (1-5378)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1985 CATAATCATAATCATAAT 2002

RESULT 324
AF068065
LOCUS       AF068065               5499 bp    DNA     linear   INV 27-DEC-1998
DEFINITION  Cryptosporidium parvum GP900 gene, complete cds.
ACCESSION   AF068065
VERSION     AF068065.1  GI:4063041
KEYWORDS    .
SOURCE      Cryptosporidium parvum
  ORGANISM  Cryptosporidium parvum
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
            Cryptosporidiidae; Cryptosporidium.
REFERENCE   1  (bases 1 to 5499)
  AUTHORS   Barnes,D.A., Bonnin,A., Huang,J.X., Gousset,L., Wu,J., Gut,J.,
            Doyle,P., Dubremetz,J.F., Ward,H. and Petersen,C.
  TITLE     A novel multi-domain mucin-like glycoprotein of Cryptosporidium
            parvum mediates invasion
  JOURNAL   Mol. Biochem. Parasitol. 96 (1-2), 93-110 (1998)
   PUBMED   9851610
REFERENCE   2  (bases 1 to 5499)
  AUTHORS   Petersen,C., Barnes,D.A. and Gousset,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAY-1998) Medicine, University of California, San
            Francisco, SFGH, Bldg. 30, Rm. 405, 1001 Potrero Ave., San
            Francisco, CA 94110, USA
FEATURES             Location/Qualifiers
     source          1. .5499
                     /organism="Cryptosporidium parvum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5807"
                     /country="USA: Iowa"
     CDS             1. .5499
                     /note="mucin-like glycoprotein"
                     /codon_start=1
                     /product="GP900"
                     /protein_id="AAC98153.1"
                     /db_xref="GI:4063042"
                     /translation="MVNIKVSSSAIALVAVIMNPLFSLAFKSSNRLEMRIESSGAVSN
                     EKFVIPSLPSDLDPTTFLLIDSTGKKFSPYTGKHADASTTSSAYSAPFELDVSGVPIE
                     PNTRRMVDPVSLMLFDNSTGVMYDPNTNSILEGSIAGIRSESCIVSELNFTSTTGFTT
                     DTSMNWPVSITSGELKDPNKQATISGSRSCGWKQGYSIDSSTGFRVDSITGLPTDPYP
                     NCPFNPVTGNLVSRSTGKTIPNTYAGVYRSNETKTTEPSANTNFLLVDPKINAPCNSE
                     NSFEQVQIFDMGSKVYIPYTKCVGVKHTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
                     TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTKKPTTTTTTTTTTTTTTTTTT
                     TTTTTTTTTTTTTTTTTTTTTTTTTTTTKKPTTTTTTTTTTTTTTTTTTTTTTTTKK
                     PTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATTTTKKPTTTTTTTTTTTTKKPT
                     TTTTATTTTTTSETESVIKPDEWCWLEKNGECEAKGATYVGVIGKDGRIENGMAFTMI
                     PNDDTHVRFRFKVKDVGNTISVRCGKGAGKLEFPDRSLDFTIPPVAGHNSCSIIVGVS
                     GGGKIHVSPYGSKDVSLISAPIQPCELFNEVYCDTCTAKYGAIHSGYQTSADFVTTTT
                     AKPTTTTTGAPGQPTTTTTGSPSKPTTTTTTKATTTTTILNPIITTTTQKPTTTTTTK
                     VPGKPPIATTTTTLKPIVTTTTTKATTTTTTTVPTTTTTTKRDEMTTTTTPLPDIGDI
                     EITPIPIEKMLDKYTRMIYDYNSGLLLDSNDEPIPGSQAGQIADTSNLFPVQTHKSTG
                     LPIDPMVGLPFDPKSGNLVHPYTNQTMSGLSVSYLAAKNLTVDTDETYGLPIDTLTGY
                     PLDPVSLIPFNPETGELFDPISDEIMNGTIAGIVSGISASESLLSQKSAPIDPATNMV
                     VGEFGGLLNPATGVMIPGSLGPSEQTPFSPEIEDGGIIPPEVAAANADKFKLSIPPSV
                     PESIPEKDQKIDSISELMYDIESGRLIGQVSKRPIPGSIAGDLNPIMKTPTQTDSVTG
```

```
                     KPIDPTTGLPFNPPTGHLINPTNNNTMDSSFAGAYKYAVSNGIKTDNVYGLPVDEITG
                     LPKDPVSDIPFNSTTGELVDPSTGKPINNYTAGIVSGKRGLPPIEDENGNLFDPSTKL
                     PIDGNNQLVNPETNSTVSGSTSGSTKPKPGIPVNGGGVVPDEEAKDQADKGKDGLIVP
                     PTNSINKDPVTNTQYSNTTGNIINPETGKVIPGSLPGSLNYPSFNTPQQTDEITGKPV
                     DTVTGLPYDPSTGEIIDPATKLPIPGSVAGDEILTEVLNITTDEVTGLPIDLETGLPR
                     DPVSGLPQLPNGTLVDPSNKKPIPGSHSGFINGTSGEQSHEKDPSTGKPLDPNTGLPF
                     DEDSGSLINPETGDKLQGSHSGTFMPVPGKPQGENGGIMTPEQILEALNKLPTSNEVN
                     ISPRPSSDAVPDRPTNTWWNKISGQTFQVDGKKTIPGSAASVIHTALGTPTQTDPTTG
                     LPSDPSTGLPFIPGFNVLVDPQTGEQIKGSVPYVSLYVKEKNIVTEAAYGLPVDPKTG
                     FPIDPISYLPFAKNGELIDPISGKYFSGSIAGFISGKAGSQSKSSDESGNPIDPSTNM
                     PYDPKTGKLIDPESGIAIDNSVSGVFATVPGTAAPKKGGVIPESVAAEAAKKYFAANV
                     EGEGEGEEVPPPPESSSNIAIQAAGGASAAVGLVAAVGAWYASRNRQEGEDDDDYQMD
                     LKQNMKKKRKKRVMKQQMKLLLQLSVIHHSGTNLKRRKDFSNSKKFRI"
     misc_feature    1. .909
                     /note="domain 1; unclassified site"
     misc_feature    910. .1557
                     /note="domain 2; unclassified site"
     misc_feature    1558. .2034
                     /note="domain 3; unclassified site"
     misc_feature    2035. .2370
                     /note="domain 4; unclassified site"
     misc_feature    2371. .5496
                     /note="domain 5; unclassified site"
     misc_feature    5233. .5304
                     /note="putative; transmembrane-region site"
     misc_feature    5305. .5496
                     /note="putative cytoplasmic region; unclassified site"
ORIGIN

Alignment Scores:
Pred. No.:              1.21e+03        Length:         5499
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-858-332G-28 (1-6) x AF068065 (1-5499)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1227 CACAACCACAACCACAAC 1244

RESULT 325
BT010107
LOCUS       BT010107                5501 bp    mRNA    linear   INV 12-AUG-2003
DEFINITION  Drosophila melanogaster GH05223 full insert cDNA.
ACCESSION   BT010107
VERSION     BT010107.1  GI:33589618
KEYWORDS    FLI_CDNA.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 5501)
  AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J.,
            Champe,M., Chavez,C., Dorsett,V., Dresnek,D., Farfan,D., Frise,E.,
            George,R., Gonzalez,M., Guarin,H., Kronmiller,B., Li,P., Liao,G.,
            Miranda,A., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
            Patel,S., Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Rubin,G.M.
            and Celniker,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-AUG-2003) Berkeley Drosophila Genome Project,
            Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, USA
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
             Lawrence Berkeley National Laboratory
            Berkeley, CA 94720
             This clone was sequenced as part of a high-throughput process to
            sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
            Science 2000). The sequence has been subjected to integrity checks
            for sequence accuracy, presence of a polyA tail and contiguity
            within 100 kb in the genome. Thus we believe the sequence to
            reflect accurately this particular cDNA clone. However, there are
            artifacts associated with the generation of cDNA clones that may
            have not been detected in our initial analyses such as internal
            priming, priming from contaminating genomic DNA, retained introns
            due to reverse transcription of unspliced precursor RNAs, and
            reverse transcriptase errors that result in single base changes.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our Web site
            (http://fruitfly.berkeley.edu) or send email to
            cdna@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1. .5501
                     /organism="Drosophila melanogaster"
                     /mol_type="mRNA"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
     gene            1. .5501
                     /gene="ia2"
                     /db_xref="FLYBASE:FBgn0031294"
     CDS             326. .4189
                     /gene="ia2"
                     /note="Longest ORF"
                     /codon_start=1
                     /product="GH05223p"
                     /protein_id="AAQ22576.1"
                     /db_xref="GI:33589619"
                     /db_xref="FLYBASE:FBgn0031294"
                     /translation="MARNVQQRQQHPCTTTEEPKHQRARGTAGHQESATGECQQFWSK
                     QRPPFDQQQQQEDQDPEQDPQQHPSTPQESVHHHRSPRQHELLHHQRKVPPTLPARGR
                     HLISEGARRREVRGCGAGGGGLLCCSLDALSRPAKLLQFGVLLLILLLLATHGDLAQAE
                     GNVGCQFVRTLCIPHSEVCYDDNIFGKCIPTTGVDVEDIEKTPLTEDQSRVLATMLEE
                     LQGAGLGWDHPYVQCRIQGSLFSLQRQQQLPPNLCANLAPTPPEFGDPASALAYVRFT
                     PPEPEAELDYYEQPAGPAQFYPALRKKQVRESDADDMYLNRMLQDRRRLRHDPSELEH
                     FGKMDGHGQSTQLDAPSIMDAFLDTERQRIEREQQLAAAEQDADRRAEQNRLELYQIL
                     AASEPDPQPYQRKPAAQPNAMDQLEAIVEQQQQRELKEQQEQAKAPVYVPPEEVNESS
                     ELYFPDNFAPFKRARGRSRGGLAEEVEDEQPKRSFFRELAGEQGLVQEPPLRFRHVET
                     EVDATPELVKRRPASFKQLDPYDVSLQQQEMAFESGLLRNSLTPLEEEAMLASNSFPR
                     QAKSQRVYTEGGLLLMPQDEQDTDGMQADEPENVKQSLLANMLGFARHERLDVKKPGP
                     LLGPPSLGSAELSNQLETEKAHKQKGHGNKEVLPAHIKGNDEDDAHKKKKVMKQQHSA
                     EDHAPHTVDTEYVHVFVKNPIDSWNDGQRIMKELEQILHMQGYFSYLTVQQHEVSFRV
                     NSNNPERKTAGDVARTINENRGVKNNIQRRVGFYVLHAGVGDVIKDLQDPSVSSSRIE
                     LAEQGPDVTHIMAYMFAGAGAAAVIVIFVTLILIKRHDRKRDKLGGLQSGIAGAETCS
                     KDYQELCRARMAGKAGGGGNSTGGAAGGGSNEPAPSGRITSLSKENEGRPPSSRSSTS
                     SWSEEPALTNMDISTGHMVLSYMEDHLRNKGRLQREWEALCRYEAEPSAREAASQPQC
                     AGLNRPGAPLPYDHSRVVLNHLANAEGLDYVNASTITDHDPRAPAYVAAQGPLPSTLA
                     HFWQMIWEQGAVVIVALCRLQENGEVACARYWPEEGAEVYHIYEVHLVSEHIWCDDYL
                     VRSFYLKNLRTSETRTVTQFHFLSWPHMGVPAQAKALLDFRRKVNKSYRGRRSCPIVV
                     HGSAGAGRTGVYILLDLVLERMNKGAREIDIAATLEHLRDQRAGVVATRQQFEFVLMA
                     VAEEVHAILKALPANTSGEKRELDKDPVVGGGSSSTTTKEPLKEDKAQEAAEEEAPTS
                     SSKAAAAAAKKEKEGITGGGWSFACR"
ORIGIN

Alignment Scores:
Pred. No.:              1.21e+03        Length:         5501
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-858-332G-28 (1-6) x BT010107 (1-5501)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       5205 CATAACCATAATCATAAC 5222

RESULT 326
AR097040
LOCUS       AR097040                5511 bp    DNA     linear   PAT 14-FEB-2001
DEFINITION  Sequence 2 from patent US 6071518.
ACCESSION   AR097040
```

```
VERSION     AR097040.1  GI:12805770
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5511)
  AUTHORS   Petersen,C.
  TITLE     GP900 glycoprotein and fragments for treatment and
            detection/diagnosis of cryptosporidium
  JOURNAL   Patent: US 6071518-A 2 06-JUN-2000;
FEATURES             Location/Qualifiers
     source          1. .5511
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

Alignment Scores:
Pred. No.:              1.21e+03        Length:         5511
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x AR097040 (1-5511)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1242 CACAACCACAACCACAAC 1259

RESULT 327
AR279259
LOCUS       AR279259               5511 bp    DNA     linear   PAT 10-APR-2003
DEFINITION  Sequence 2 from patent US 6514697.
ACCESSION   AR279259
VERSION     AR279259.1  GI:29713987
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5511)
  AUTHORS   Petersen,C., Barnes,D.A., Nelson,R.C. and Gut,J.
  TITLE     Methods for detection of Crytosporidium species and isolates and
            for diagnosis of Cryptosporidium infections
  JOURNAL   Patent: US 6514697-A 2 04-FEB-2003;
            The Regents of the University of California; Oakland, CA
FEATURES             Location/Qualifiers
     source          1. .5511
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              1.21e+03        Length:         5511
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x AR279259 (1-5511)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1242 CACAACCACAACCACAAC 1259

RESULT 328
AY162292/c
LOCUS       AY162292               5513 bp    mRNA    linear   INV 23-SEP-2004
DEFINITION  Aplysia californica capsulin mRNA, complete cds.
ACCESSION   AY162292
VERSION     AY162292.2  GI:31088939
KEYWORDS    .
SOURCE      Aplysia californica (California sea hare)
  ORGANISM  Aplysia californica
            Eukaryota; Metazoa; Mollusca; Gastropoda; Orthogastropoda;
            Apogastropoda; Heterobranchia; Euthyneura; Opisthobranchia;
            Anaspidea; Aplysioidea; Aplysiidae; Aplysia.
REFERENCE   1  (bases 1 to 5513)
  AUTHORS   Cummins,S.F., Nichols,A.E., Amare,A., Hummon,A.B., Sweedler,J.V.
            and Nagle,G.T.
  TITLE     Characterization of Aplysia enticin and temptin, two novel
            water-borne protein pheromones that act in concert with attractin
            to stimulate mate attraction
  JOURNAL   J. Biol. Chem. 279 (24), 25614-25622 (2004)
   PUBMED   15054104
REFERENCE   2  (bases 1 to 5513)
  AUTHORS   Nagle,G.T. and Cummins,S.F.
  TITLE     Characterization of novel pheromones in Aplysia californica
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 5513)
  AUTHORS   Nagle,G.T. and Cummins,S.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-OCT-2002) Anatomy and Neurosciences, University of
            Texas Medical Branch, 301 University Boulevard, Galveston, TX
            77555, USA
REFERENCE   4  (bases 1 to 5513)
  AUTHORS   Nagle,G.T. and Cummins,S.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAY-2003) Anatomy and Neurosciences, University of
            Texas Medical Branch, 301 University Boulevard, Galveston, TX
            77555, USA
  REMARK    Sequence update by submitter
COMMENT     On May 28, 2003 this sequence version replaced gi:26892025.
FEATURES             Location/Qualifiers
     source          1. .5513
                     /organism="Aplysia californica"
                     /mol_type="mRNA"
                     /db_xref="taxon:6500"
                     /clone="69"
                     /tissue_type="albumen gland"
     CDS             4. .5376
                     /codon_start=1
                     /product="capsulin"
                     /protein_id="AAN83924.2"
                     /db_xref="GI:31088940"
                     /translation="MMYWVLFSALLALPGVSAQSLKCAFCSSDDIEGLDLVCRASMLC
                     DQEENWGAYCYESSTVQKANISEYENKCANNTLILCEQTDSNSLTVSALKVSSNESCP
                     ANSTRQELNDNSSQSVCCATPTLESSDTGTPSRRKRSPRHRGKKIRRKDKDDGGDDDD
                     TAAATTVAPADDTKCERVFVKHENRKFFPGPVVSATLMNKKVHIAGACVDEDDYKIKG
                     NRLKLIRPCAGLFKVCYELPKTTAPATTLAPEPKCRKVYVDHAEEESVGGEIQSLDLV
                     RSIGHGGYGKGGYSLCQKGETFSFRGDKIIVKGACKGVFQVCYLETVAAATTVAPEPS
                     CQRLVVNHPSEQTLEGGNIVSVTVVKLLGAKSGGYGSYESPVSCEQDADDGYSFEGST
                     FTVTGNCKAVFRVCVAEPTQAPATTAAPEIKCRTYYVLKDTDEYANGKITSMKLQKVT
                     SYGYCHKKADYGFSGSRFYVKNGCRGQFKICIEKPVAPPPPPTTTILPATTTVAPEPV
                     CEKKYIYQDKDGSVDGDIVSFKLIKSLKRPQPQGYGSYGYDYCRRGYNYKFSGSSFSV
                     RDCPGVYEICYNAVITEAPATTTPNPEVCKKFVAKHGLKEDLPGQITSVELLSNVEEK
                     AGSHGYGYGYPAPSYNYCKKWTHYFFYDSQFTVTSDCAGIFKVCYNPAPPTQAPARTT
                     AEPNTVCRKEIVDDKKKIRVNGVIVSVKLIKATPKASGYGYQNPDYCYENKDYFYKDD
                     TLDVRSCPGVYNICYEERKTTAKATTTVIPEKCKRVSSSTTFAGEITSVEILKANPKG
                     GYGHGYDYNSCYKDVDYFYSGNAFTLVPTCKSTFKVCYQEPQPTTAPATTLAPEKQCK
                     KVNLDKYGKESFGGKIVSVQYVPADYRSYDPHGPSCREDKDYVINGNTLETVTCAGWF
                     KVCYEQSVTEAPATTTVTPEKCERVSSYKGVKKNFGGNVVSAKLISSLGAGGSGGYGG
                     GYYHCTENVDYFFQGSQFVVLANCKAVFKVCYEDTATEAPATTLAPEPECENVHIVKK
                     GNREFGGDIISVEVVKGIVIRSNSYGYGPASYGYGSCYEGSDYSFSGSVLSVRSCSGV
                     FKVCYERPQTVAPATTVSPDICKKVVSYYGVKKSFNGEIVSVKLGPTLGVKSPGNGYG
                     YGNSYGSAQGRGYNQGSGYGYGSRYPSLGNSQGYGYPLSQGYGYNSGRQGNGYGHGYS
                     TCREGLTYYYHGNEFVALSQCRAVFTVCYRETATTVAPATTVAPEPKCRKFKAYAGDD
                     KEFDGDIVSVRLLKAFRKPGYGYREPSCKEGKDYVSSGNVLHVLQCAGFYEVCYEERF
                     TTLPATTVAPETECRKFIATHKLEETVGGRIVSADLVQRLTKQTGYGPNKGIGYGPGR
                     GIGYGPSHGIGYGPSHGIGYGPKGSRSYAPGYGQRIGYGYKSSYDQNIGYGYGSSNSR
                     GSGYGHGNGYGQGYGYGYGGGSQNGGFGQRTGYGYGYDECKEGVGYFFNDNRFTVTPL
                     CNAVFKVCYELPVPTTAPATTVGPEPVCEKVNIYGQERKEYDGDIVRVRLIKVNGPHG
                     PPGGYGYSRPCREGIDYLFQGNVFNVKNCNALFFVCYEPKITTAPATTTETPVTCKKF
                     KVSYGTKEKFDGKITSAALISKFGVHGGYGGGYGYGGHPQCSRGYGYDYNNNVFTAQA
                     QCSGVYKVCFENPATTAPATTVAPAEPTCEKLSVWFNEEKTFEGSRIVSFKLVRLFNR
```

YKKCKKDQYGVKGDEENTFVVKGCSGVFQVCYEEQTAPATTTEAPKPEPRKPRKKKFP
FSIGKY"
ORIGIN

Alignment Scores:
Pred. No.: 1.21e+03 Length: 5513
Score: 42.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 2 Gaps: 0

US-09-858-332G-28 (1-6) x AY162292 (1-5513)

Qy 1 HisAsnHisAsnHisAsn 6
||||||||||||||||||
Db 4423 CATAACCATAACCATAAC 4406

RESULT 329
AX346923/c
LOCUS AX346923 5542 bp DNA linear PAT 01-FEB-2002
DEFINITION Sequence 1994 from Patent WO0200928.
ACCESSION AX346923
VERSION AX346923.1 GI:18494809
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1
AUTHORS Olek,A., Piepenbrock,C. and Berlin,K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patent: WO 0200928-A 1994 03-JAN-2002;
Epigenomics AG (DE)
FEATURES Location/Qualifiers
source 1. .5542
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

Alignment Scores:
Pred. No.: 1.22e+03 Length: 5542
Score: 42.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-858-332G-28 (1-6) x AX346923 (1-5542)

Qy 1 HisAsnHisAsnHisAsn 6
||||||||||||||||||
Db 2870 CACAATCACAATCATAAT 2853

RESULT 330
CQ807300/c
LOCUS CQ807300 5559 bp DNA linear PAT 10-MAY-2004
DEFINITION Sequence 750 from Patent WO2004035803.
ACCESSION CQ807300
VERSION CQ807300.1 GI:47112682
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1
AUTHORS Foekens,J., Harbeck,N., Koenig,T., Maier,S., Martens,J., Model,F., Nimmrich,I., Rujan,T., Schmitt,A., Schmitt,M., Look,M.P. and Marx,A.
TITLE Method and nucleic acids for the improved treatment of breast cell proliferative disorders
JOURNAL Patent: WO 2004035803-A 750 29-APR-2004;
Epigenomics AG (DE)
FEATURES Location/Qualifiers
source 1. .5559
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

Alignment Scores:
Pred. No.: 1.22e+03 Length: 5559
Score: 42.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-858-332G-28 (1-6) x CQ807300 (1-5559)

Qy 1 HisAsnHisAsnHisAsn 6
||||||||||||||||||
Db 1979 CACAACCATAACCACAAC 1962

RESULT 331
AX251238/c
LOCUS AX251238 5571 bp DNA linear PAT 05-OCT-2001
DEFINITION Sequence 206 from Patent WO0168912.
ACCESSION AX251238
VERSION AX251238.1 GI:15984661
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1
AUTHORS Olek,A., Piepenbrock,C. and Berlin,K.
TITLE Diagnosis of diseases associated with tumor suppressor genes and oncogenes
JOURNAL Patent: WO 0168912-A 206 20-SEP-2001;
Epigenomics AG (DE)
FEATURES Location/Qualifiers
source 1. .5571
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

Alignment Scores:
Pred. No.: 1.22e+03 Length: 5571
Score: 42.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-858-332G-28 (1-6) x AX251238 (1-5571)

Qy 1 HisAsnHisAsnHisAsn 6
||||||||||||||||||
Db 306 CACAACCACAACCACAAC 289

RESULT 332
AX251941/c
LOCUS AX251941 5571 bp DNA linear PAT 05-OCT-2001
DEFINITION Sequence 202 from Patent WO0168911.
ACCESSION AX251941
VERSION AX251941.1 GI:15985296
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1
AUTHORS Olek,A., Piepenbrock,C. and Berlin,K.
TITLE Diagnosis of diseases associated with the cell cycle

```
  JOURNAL   Patent: WO 0168911-A 202 20-SEP-2001;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .5571
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

Alignment Scores:
Pred. No.:              1.22e+03       Length:         5571
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     6              Gaps:           0

US-09-858-332G-28 (1-6) x AX251941 (1-5571)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        306 CACAACCACAACCACAAC 289

RESULT 333
AX344343/c
LOCUS       AX344343               5571 bp    DNA     linear   PAT 01-FEB-2002
DEFINITION  Sequence 190 from Patent WO0200926.
ACCESSION   AX344343
VERSION     AX344343.1  GI:18492231
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with signal transduction
  JOURNAL   Patent: WO 0200926-A 190 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .5571
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

Alignment Scores:
Pred. No.:              1.22e+03       Length:         5571
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     6              Gaps:           0

US-09-858-332G-28 (1-6) x AX344343 (1-5571)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        306 CACAACCACAACCACAAC 289

RESULT 334
AX348740/c
LOCUS       AX348740               5571 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 198 from Patent WO0202807.
ACCESSION   AX348740
VERSION     AX348740.1  GI:18614775
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with cell signalling
  JOURNAL   Patent: WO 0202807-A 198 10-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .5571
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

Alignment Scores:
Pred. No.:              1.22e+03       Length:         5571
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     6              Gaps:           0

US-09-858-332G-28 (1-6) x AX348740 (1-5571)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        306 CACAACCACAACCACAAC 289

RESULT 335
AX345106/c
LOCUS       AX345106               5621 bp    DNA     linear   PAT 01-FEB-2002
DEFINITION  Sequence 177 from Patent WO0200928.
ACCESSION   AX345106
VERSION     AX345106.1  GI:18492992
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with the immune system
  JOURNAL   Patent: WO 0200928-A 177 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .5621
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

Alignment Scores:
Pred. No.:              1.23e+03       Length:         5621
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     6              Gaps:           0

US-09-858-332G-28 (1-6) x AX345106 (1-5621)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        844 CACAACCACAATCATAAC 827

RESULT 336
AY122262
LOCUS       AY122262               5687 bp    mRNA    linear   INV 19-JUN-2002
DEFINITION  Drosophila melanogaster SD13096 full insert cDNA.
ACCESSION   AY122262
VERSION     AY122262.1  GI:21483597
KEYWORDS    FLI_CDNA.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
```

```
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 5687)
  AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J.,
            Champe,M., Chavez,C., Dorsett,V., Dresnek,D., Farfan,D., Frise,E.,
            George,R., Gonzalez,M., Guarin,H., Kronmiller,B., Li,P., Liao,G.,
            Miranda,A., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
            Patel,S., Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Rubin,G.M.
            and Celniker,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2002) Berkeley Drosophila Genome Project,
            Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, USA
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
             Lawrence Berkeley National Laboratory
            Berkeley, CA 94720
             This clone was sequenced as part of a high-throughput process to
            sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
            Science 2000). The sequence has been subjected to integrity checks
            for sequence accuracy, presence of a polyA tail and contiguity
            within 100 kb in the genome. Thus we believe the sequence to
            reflect accurately this particular cDNA clone. However, there are
            artifacts associated with the generation of cDNA clones that may
            have not been detected in our initial analyses such as internal
            priming, priming from contaminating genomic DNA, retained introns
            due to reverse transcription of unspliced precursor RNAs, and
            reverse transcriptase errors that result in single base changes.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our Web site
            (http://fruitfly.berkeley.edu) or send email to
            cdna@fruitfly.berkeley.edu.
FEATURES             Location/Qualifiers
     source          1. .5687
                     /organism="Drosophila melanogaster"
                     /mol_type="mRNA"
                     /db_xref="taxon:7227"
     gene            1. .5687
                     /gene="CG10231"
                     /db_xref="FLYBASE:FBgn0032686"
     CDS             411. .5045
                     /gene="CG10231"
                     /note="Longest ORF"
                     /codon_start=1
                     /product="SD13096p"
                     /protein_id="AAM52774.1"
                     /db_xref="GI:21483598"
                     /db_xref="FLYBASE:FBgn0032686"
                     /translation="MKVTQSEENTRNTSDRSKSVQTNRKFDNFNWLLSCGLLAAVKST
                     KLIPQLPRQQQPKKYNLRYAAELLKFDKIQFIKTEGLTFALLAGPMGQAASMCRFRGC
                     RYKNKNKSSKQQQQQQQQQQQQQQQQQHQQQQQQTPSHSPQIQHHSEIIPATTGLHLR
                     SIEEPATTPLQFQPTGRMNTEQGGTGYGGYGSSEHSLLIATRHAGVPLPLAQHQPLPA
                     HYQPLNHSGAAPPSSSNGSSSSGGGVQTSATPQQQQQYQVQQPYQYQYQHHYHHQANS
                     PQHHRPYDPEHARMEAWLDENQEFVQDYFIRKATRQTVDAWLVSHATSAGNDVVSSTS
                     PTHANGQTSSSRGGSGATTPVRKISAHEFERGGLLKPIVNTIDGTPTFLSIGPPMDNG
                     SVGGSCSNLQNVGGVVAGQYQYNHQQHHHNHAHLHHSQHSHYQAGGAVGSSSLGSTGG
                     ASGAGGAPSLGGSGGAGNGHQYPYYHCHQRPQRLSRNELKQLDEKELIFELVKDICNE
                     LEVRTLCHKILQNVSILLNADRGSLFLVQGRCNGPDGLKKCLVSKLFDVCPRSTVEEM
                     EQQDEVRVAWGTGIAGHVAESGEPVNIPDAYQDERFNCEIDSLTGYRTKALLCMPIKD
                     SSGDVIGVAQVINKMNGECFSEIDEKVFSSYLQFCGIGLRNAQLYEKSQLEIKRNQVL
                     LDLARMIFEEQSTIEHMVFRILTHMQSLIQCQRVQILLVHEADKGSFSRVFDFEANDL
                     SEEEATSRTSPYESRFPINIGITGHVATTGETVNVPNAYEDDRFDASVDENSCFKHRS
                     ILCMAIKNSLGQIIGVIQLINKFNELDFTKNDENFVEAFAIFCGMGIHNTHMYEKAIV
                     AMAKQSVTLEVLSYHASATMDEAHRLRRLRVPSAVHFRLHDFKFDDIHFEDDDTLKAC
                     LRMFLDLDFVERFHIDYEVLCRWLLSVKKNYRNVTYHNWRHAFNVAQMMFAILTTTQW
                     WKIFGEIECLALIIGCLCHDLDHRGTNNSFQIKASSPLAQLYSTSTMEHHHFDQCLMI
                     LNSPGNQILANLSSDDYCRVIRVLEDAILSTDLAVYFKKRGPFLESVSQPTSYWVAEE
                     PRALLRAMSMTVCDLSAITKPWEIEKRVADLVSSEFFEQGDMEKQELNITPIDIMNRE
                     KEDELPMMQVNFIDSICLPIYEAFATLSDKLEPLVEGVRDNRGHWIDLADVVKTKTSQ
                     DQEPEEEQQQQNVISNGDCKAMSDDDVAASEAEVAVDSPSEKASVNGSNVANNSSNTN
                     KKIAVASHPTSTQPSDDDNDADADVDDVDEQAAEENGHDAEVDEASCRSNSTCSSSTA
                     SSRLSTPPPTGEDDSTPVSPLKTLQAKLVAANLNALQRQTSNQAQTQKQRCKSCDHSR
                     SGLQVRKTSSLRGAQELDLDSKTRNGTHAALCKSTPVINNHSHHHNHSHSHNHNHNHH
                     HHHHHHSHHNHSQHGIGIGSASIGGSGLISLTTPLLAMDSDRIPKIVGKIGNLDGLPF
                     ANGIGGPQNGHGLPFGSYQHHHHHQHHHHLLARRHSETNSNGATAMAVEK"
ORIGIN

Alignment Scores:
Pred. No.:                1.25e+03        Length:          5687
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       2               Gaps:            0

US-09-858-332G-28 (1-6) x AY122262 (1-5687)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       4695 CACAACCACAACCACAAC 4712

RESULT 337
AB167716/c
LOCUS       AB167716                5728 bp    DNA     linear   PLN 01-SEP-2004
DEFINITION  Rhizopus oryzae ldhB gene for lactate dehydrogenase, complete cds,
            strain: NBRC 4747.
ACCESSION   AB167716
VERSION     AB167716.1  GI:51775930
KEYWORDS    .
SOURCE      Rhizopus oryzae
  ORGANISM  Rhizopus oryzae
            Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;
            Rhizopus.
REFERENCE   1
  AUTHORS   Saito,K., Saito,A., Ohnishi,M. and Oda,Y.
  TITLE     Genetic diversity in Rhizopus oryzae strains as revealed by the
            sequence of lactate dehydrogenase genes
  JOURNAL   Arch. Microbiol. 182 (1), 30-36 (2004)
   PUBMED   15278242
REFERENCE   2  (bases 1 to 5728)
  AUTHORS   Saito,K., Saito,A., Ohnishi,M. and Oda,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAR-2004) Katsuichi Saito, National Agriculture and
            Bio-oriented Research Organaization, National Agricultural Research
            Center for Hokkaido Region; Shinsei, Memuro, Kasai, Hokkaido
            082-0071, Japan (E-mail:k.saito@affrc.go.jp, Tel:81-155-62-9280,
            Fax:81-155-62-9281)
FEATURES             Location/Qualifiers
     source          1. .5728
                     /organism="Rhizopus oryzae"
                     /mol_type="genomic DNA"
                     /strain="NBRC 4747"
                     /db_xref="taxon:64495"
     CAAT_signal     2504. .2507
     TATA_signal     2563. .2567
     gene            2676. .3584
                     /gene="ldhB"
     CDS             2676. .3584
                     /gene="ldhB"
                     /EC_number="1.1.1.27"
                     /codon_start=1
                     /product="lactate dehydrogenase"
                     /protein_id="BAD38977.1"
                     /db_xref="GI:51775931"
                     /translation="MVLHSKVAIIGAGAVGASTAYALMFKNICTEIIIVDVNPDIVQA
                     QVLDLADAASVSNTPIRAGSAEEAGQSDIIVITAGAKQKEGEPRTKLIERNYRVLKNI
                     IGGMQPIRSDAIILVVVNPVDILTHIAQTLSGLAPNQVIGSGTYLDTTRLCVHLGDIF
                     DVNPQSIHAFVLGEHGDSQMIAWEAASIGGQPLTSFPEFAELDKKAISKAISGKAMEI
                     IRLKGATFYGIGACAADLVHTIMLNRKSVHPVSVYVEKYGVTFSMPAKLGWRGVEKIY
                     EVPLTEEEEALLLKSVEALKAVEYLS"
     polyA_signal    3599. .3604
ORIGIN

Alignment Scores:
Pred. No.:                1.26e+03        Length:          5728
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
```

```
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     15            Gaps:          0

US-09-858-332G-28 (1-6) x AB167716 (1-5728)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       4991 CACAATCACAACCACAAC 4974

RESULT 338
AF355378
LOCUS       AF355378               5772 bp    DNA     linear   PLN 01-MAY-2001
DEFINITION  Zea mays bz-m2(Ac) mutable allele, genomic sequence.
ACCESSION   AF355378
VERSION     AF355378.1  GI:13898794
KEYWORDS    .
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 5772)
  AUTHORS   Fu,H. and Dooner,H.K.
  TITLE     The Ac element in the bz-m2(Ac) allele from maize
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 5772)
  AUTHORS   Fu,H. and Dooner,H.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2001) Waksman Institute, Rutgers University,
            Piscataway, NJ 08854, USA
FEATURES             Location/Qualifiers
     source          1. .5772
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4577"
                     /chromosome="9"
                     /map="9S"
     source          525. .5089
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4577"
                     /chromosome="9"
                     /map="9S"
                     /transposon="Activator (Ac)"
                     /note="Ac inserted into the Bz-McC allele"
     gene            <1. .>5772
                     /gene="bz"
                     /allele="bz-m2(Ac)"
     misc_feature    517. .524
                     /gene="bz"
                     /allele="bz-m2(Ac)"
                     /note="duplicated integration site sequence"
     misc_feature    5090. .5097
                     /gene="bz"
                     /allele="bz-m2(Ac)"
                     /note="duplicated integration site sequence"
ORIGIN

Alignment Scores:
Pred. No.:              1.27e+03      Length:        5772
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     15            Gaps:          0

US-09-858-332G-28 (1-6) x AF355378 (1-5772)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1944 CACAACCACAACCACAAC 1961

RESULT 339
AX345465/c
LOCUS       AX345465               5826 bp    DNA     linear   PAT 01-FEB-2002
DEFINITION  Sequence 536 from Patent WO0200928.
ACCESSION   AX345465
VERSION     AX345465.1  GI:18493351
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with the immune system
  JOURNAL   Patent: WO 0200928-A 536 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .5826
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

Alignment Scores:
Pred. No.:              1.28e+03      Length:        5826
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-28 (1-6) x AX345465 (1-5826)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        233 CATAACCATAACCATAAC 216

RESULT 340
AR096206
LOCUS       AR096206               6000 bp    DNA     linear   PAT 08-SEP-2000
DEFINITION  Sequence 9 from patent US 6005092.
ACCESSION   AR096206
VERSION     AR096206.1  GI:10024799
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 6000)
  AUTHORS   Shoseyov,O. and Shani,Z.
  TITLE     Arabidopsis thaliana endo-1,4-.beta.-glucanase gene and promoter
  JOURNAL   Patent: US 6005092-A 9 21-DEC-1999;
FEATURES             Location/Qualifiers
     source          1. .6000
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

Alignment Scores:
Pred. No.:              1.31e+03      Length:        6000
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-28 (1-6) x AR096206 (1-6000)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        996 CATAATCACAACCACAAC 1013

RESULT 341
BD080275
```

```
LOCUS       BD080275                6000 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Transgenic plant with altered morphology and
            endo-1,4-beta-glucanase gene, promoter and protein isolated from
            Arabidopsis thaliana.
ACCESSION   BD080275
VERSION     BD080275.1  GI:22625878
KEYWORDS    JP 2001512672-A/6.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 6000)
  AUTHORS   Shoseyov,O., Shani,Z. and Shpiegl,E.
  TITLE     Transgenic plant with altered morphology and
            endo-1,4-beta-glucanase gene, promoter and protein isolated from
            Arabidopsis thaliana
  JOURNAL   Patent: JP 2001512672-A 6 28-AUG-2001;
            YISSUM RESEARCH DEVELOPMENT CO OF THE HEBREW UNIVERSITY OF
            JERUSALEM SECRETARY OF THE DEPARTMENT OF HEAL
COMMENT     OS   Nucleic acid
            PN   JP 2001512672-A/6
            PD   28-AUG-2001
            PF   26-JUL-1998 JP 2000506315
            PR   27-JUL-1997 IL       121404,13-JAN-1998 US    09/006632 PR
            13-JAN-1998 US    09/006636
            PI   ODED SHOSEYOV,ZIV SHANI,ETAI SHPIEGL
            PC   A01H5/00,C12N5/10,C12N9/42,C12N15/09,C12N5/00,C12N15/00 CC
            Transgenic plant with altered morphology and endo-1,4-beta- CC
                 glucanase gene,
            CC   promoter and protein isolated from Arabidopsis thaliana FH
            Key             Location/Qualifiers
            FT   source          1. .6000
            FT                   /organism='Nucleic acid'.
FEATURES             Location/Qualifiers
     source          1. .6000
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN

Alignment Scores:
Pred. No.:                1.31e+03        Length:          6000
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       6               Gaps:            0

US-09-858-332G-28 (1-6) x BD080275 (1-6000)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        996 CATAATCACAACCACAAC 1013

RESULT 342
AR262069
LOCUS       AR262069                6000 bp    DNA     linear   PAT 29-JAN-2003
DEFINITION  Sequence 9 from patent US 6323023.
ACCESSION   AR262069
VERSION     AR262069.1  GI:28073416
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 6000)
  AUTHORS   Shoseyov,O. and Shani,Z.
  TITLE     Vectors containing nucleic acids coding for Arabidopsis thaliana
            endo-1,4-.beta.-glucanase secretion signal peptide
  JOURNAL   Patent: US 6323023-A 9 27-NOV-2001;
            Yissum Research Development Co., Ltd.; Jerusalem;
            ILX;
FEATURES             Location/Qualifiers
     source          1. .6000
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:                1.31e+03        Length:          6000
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       6               Gaps:            0

US-09-858-332G-28 (1-6) x AR262069 (1-6000)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        996 CATAATCACAACCACAAC 1013

RESULT 343
ATX98543
LOCUS       ATX98543                6000 bp    DNA     linear   PLN 18-APR-2005
DEFINITION  A.thaliana endo-1,4-beta-glucanase gene.
ACCESSION   X98543
VERSION     X98543.1  GI:2440032
KEYWORDS    endo-1,4-beta-glucanase.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Shani,Z., Dekel,M., Tsabary,G. and Shoseyov,O.
  TITLE     Cloning and characterization of elongation specific
            endo-1,4-beta-glucanase (cel1) from Arabidopsis thaliana
  JOURNAL   Plant Mol. Biol. 34 (6), 837-842 (1997)
  PUBMED    9290636
REFERENCE   2  (bases 1 to 6000)
  AUTHORS   Shoseyov,O.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUN-1996) O. Shoseyov, Hebrew University of
            Jerusalem, Horticulture, The Faculty of Agricultural, Rehovot,
            76100, ISRAEL
FEATURES             Location/Qualifiers
     source          1. .6000
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /clone_lib="lambda EMBL3"
                     /ecotype="Columbia"
     gene            join(1666. .1876,1960. .2368,2747. .2836,2935. .3097,
                     4383. .4493,4661. .4825,4941. .5273)
                     /gene="egase"
     CDS             join(1666. .1876,1960. .2368,2747. .2836,2935. .3097,
                     4383. .4493,4661. .4825,4941. .5273)
                     /gene="egase"
                     /codon_start=1
                     /product="endo-1,4-beta-glucanase"
                     /protein_id="CAA67156.1"
                     /db_xref="GI:2440033"
                     /db_xref="GOA:O23696"
                     /db_xref="InterPro:IPR001701"
                     /db_xref="UniProt/TrEMBL:O23696"
                     /translation="MARKSLIFPVILLAVLLFSPPIYSAGHDYRDALRKSILFFEGQR
                     SGKLPPDQRLKWRRDSALRDGSSAGVNLSGGYYDAGDNIKFGFPMAFTTTMLSWSIID
                     FGKTMGPELRNAVKAVKWGTDYLLKATAIPGVVFVQVGDAYSDHNCWERPEDMDTLRT
                     VYKIDRAHPGSDVAGETAAALAAASIVFRKRDPAYSRLLLDRATRVRSRLLTDIAARI
                     VTVSTTRFVLFTVISTVTRDELLWGAAWLHKASRKRAYREFIVKNEVILKAGDTINEF
                     GWDNKHAGINVLISKEVLMGKAEYFESFKQNADGFICSILPGISHPQVQYSRGGLLVK
                     TGGSNMQHVTSLSFLLLAYSNYLSHAKKVVPCGELTASPSLLRQIAKRQVDYILGDNP
                     MGLSYMVGYGQKFPRRIHHRGSSVPSVSAHPSHIGCKEGSRYFLSPNPNPNLWVGAVV
                     GGPNVTDAFPDSRPYFQQSEPTTYINAPLVGLLGYFSAHSTWR"
     exon            <1666. .1876
                     /gene="egase"
                     /number=1
```

```
     intron          1877. .1959
                     /gene="egase"
                     /number=1
     exon            1960. .2368
                     /gene="egase"
                     /number=2
     intron          2369. .2746
                     /gene="egase"
                     /number=2
     exon            2747. .2836
                     /gene="egase"
                     /number=3
     intron          2837. .2934
                     /gene="egase"
                     /number=3
     exon            2935. .3097
                     /gene="egase"
                     /number=4
     intron          3098. .4382
                     /gene="egase"
                     /number=4
     exon            4383. .4493
                     /gene="egase"
                     /number=5
     intron          4494. .4660
                     /gene="egase"
                     /number=5
     exon            4661. .4825
                     /gene="egase"
                     /number=6
     intron          4826. .4940
                     /gene="egase"
                     /number=6
     exon            4941. .>5273
                     /gene="egase"
                     /number=7
ORIGIN

Alignment Scores:
Pred. No.:              1.31e+03        Length:         6000
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x ATX98543 (1-6000)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        996 CATAATCACAACCACAAC 1013

RESULT 344
AX344794/c
LOCUS       AX344794                6236 bp    DNA     linear   PAT 01-FEB-2002
DEFINITION  Sequence 219 from Patent WO0200927.
ACCESSION   AX344794
VERSION     AX344794.1  GI:18492680
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with development genes
  JOURNAL   Patent: WO 0200927-A 219 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .6236
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

Alignment Scores:
Pred. No.:              1.36e+03        Length:         6236
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x AX344794 (1-6236)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       4139 CATAACCATAACCATAAC 4122

RESULT 345
AX346532/c
LOCUS       AX346532                6236 bp    DNA     linear   PAT 01-FEB-2002
DEFINITION  Sequence 1603 from Patent WO0200928.
ACCESSION   AX346532
VERSION     AX346532.1  GI:18494418
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with the immune system
  JOURNAL   Patent: WO 0200928-A 1603 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .6236
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

Alignment Scores:
Pred. No.:              1.36e+03        Length:         6236
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x AX346532 (1-6236)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       4139 CATAACCATAACCATAAC 4122

RESULT 346
AY160091
LOCUS       AY160091                6289 bp    DNA     linear   INV 23-OCT-2002
DEFINITION  Dictyostelium discoideum nucleotide exchange factor RasGEF B (gefB)
            gene, complete cds.
ACCESSION   AY160091
VERSION     AY160091.1  GI:24286633
KEYWORDS    .
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 6289)
  AUTHORS   Wilkins,A., Szafranski,K., Gloeckner,G., Harrisingh,M.,
            Deenadayalan,B., Mueller,R., Eichinger,L., Noegel,A.A. and
            Insall,R.
  TITLE     The family of rasGEF genes in Dictyostelium discoideum
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 6289)
  AUTHORS   Szafranski,K.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (09-OCT-2002) Abt. Genomanalyse, Institut fuer Molekulare
            Biotechnologie, Beutenbergstrasse 11, D-07745 Jena, Germany
FEATURES             Location/Qualifiers
     source          1. .6289
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
     gene            <938. .>5882
                     /gene="gefB"
     mRNA            join(<938. .1039,1140. .1532,1681. .1825,1933. .>5882)
                     /gene="gefB"
                     /product="nucleotide exchange factor RasGEF B"
     CDS             join(938. .1039,1140. .1532,1681. .1825,1933. .5882)
                     /gene="gefB"
                     /note="GefB"
                     /codon_start=1
                     /product="nucleotide exchange factor RasGEF B"
                     /protein_id="AAN46871.1"
                     /db_xref="GI:24286634"
                     /translation="MVVILWGWNENRFNTLDQVISFDFIEENKKLYLKILENEKQQKI
                     YNISNLIGRRMDNYRDNNNNNGNGSNINNNGIIINSSMGSQYFNLNKGIIIVESDNET
                     FLIEEIDYLGKFPILKQFIKCEEFIGISTLEIISNIEKQLSNLVNLKSNTTEQTDRKY
                     KTNLIDFKESIIQLEKDCKNLLKQSNLINPSIKYSVVKSVKFLLKEIKFEFYHGDINI
                     KRIELFQQLKKFLKIHFWNNPNLILNSGSNIDMTDSNDSISSEEINSPNDLNLSLSSF
                     PSISSPSNSSYNNINTLTFSNNNNNNNNSNNNNNNNNNNNNSIINSNNNNNSNNN
                     SNISNSSNNNNNNSNNSNNISINVSKTGNNSVINSPTQSSSCYIDCDDTMSNYTDCTN
                     TTNTNSTTTTTTTTTTYGNNLNVYNNNNNNNTNNNNNNVPYKKPALKMMKSPSFKIQL
                     IQEVYLEIMDQCKKIQNYLYELLDTYLTDEEYINNNNNNNNKDISILFENNTSYCIAQ
                     EIFISLIKNSNDFLYQVQLIDLTDTRKKSKELSEVSNQLFYSTQQLFPSISDHLDFIE
                     MDIFTNNNTTTTTTTTTTTTNTTTNPSNTIKSSHSSLLSSSYSSSKDDDIVLTINKVL
                     SKLDTFSKEYTKLIGNNLSSNYYVDSGNNNESVSSPKCGLFLSHSLSTDSPMTYVKRN
                     TSSGSSASTSSYTTPVSSSASLSFSSSTQTPSSAASQHHIQQILQQQQQQQQQQQQQQ
                     QQQQPQQIQSQSQQQHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQIQSQSQQQQQLNRK
                     PHYSSISPTSMLVYNATGGKFGVVNQNNNSAESTPTFQLNTMDILNSLMGKEGQNSNT
                     NNANNPNNNNNNNNNNNNNNNNNNNNNNNNNNENKNENKNETNKSGHNYNNSSTSTLSSS
                     TTIGDLANVSSTNSPNSSTPNLLLPPHQFHNHNREYNHSHHHLLSSSSNNVMINNRSI
                     GKSFSTGNFSQPQPIMPSNGIGGSGGGGLGTTLSVGNGGIVGGLQSRSSSFEENQLQY
                     AIESFMEVYNHLTSKDKVDTSIWLEEEKEDINVYYINNTNIDDKFYRSIKYASLNQLI
                     LKLTCESNTELRFTKTFITTYRSFTTGDIFLMKLIQRYFTPNLKNIVPYQFVQKIQIP
                     IQLRVLNVLRLWIEQYPSDFQQPPQITTQSTQPIQNSTTQPQPQPQPQQPQPQLQQST
                     NLQNSTIQQPSLFKMLTAFLNSTRRNGHGQYSDLILKKFKQQKQKDRLQLPIINNGQG
                     IPKAKIFWMKYSTEFIFSQSSTDIAEQLTLLDFDSYKSIEEIELLNQAWSKPDQKTNT
                     PNIASMVNRFNSFSSFVSWAILRENDIKQRSKMMLKMIKICYALYKLSNFNGLLAGLS
                     GLMASGVYRLNHTKSLISKQYQKKFDFLCKFLDTKKSHKVYRDIIHSTCPPLIPYLGI
                     YLTDLTFIEDGNQDEIKGLINFKKREMIFNTILEIQQYQQQGYTIKPKPSVLGFLCEL
                     PHMSDKKKFEDDTYEQSILLEPKNSTKLEVMNAKK"
ORIGIN

Alignment Scores:
Pred. No.:                1.37e+03        Length:         6289
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       2               Gaps:           0

US-09-858-332G-28 (1-6) x AY160091 (1-6289)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       4846 CACAACCACAACCACAAC 4863

RESULT 347
SCYGL178W
LOCUS       SCYGL178W               6323 bp    DNA     linear   PLN 18-APR-2005
DEFINITION  S.cerevisiae chromosome VII reading frame ORF YGL178w.
ACCESSION   Z72700 Y13135
VERSION     Z72700.1  GI:1322785
KEYWORDS    .
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 6323)
  AUTHORS   Bruschi,C.V., Coglievina,M., Bertani,I., Klima,R., Zaccaria,P. and
            Delneri,D.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 6323)
  AUTHORS   MIPS.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAY-1996) Data collected by MIPS on behalf of the
            European yeast chromosome VII sequencing project. MIPS at the
            Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
            Martinsried, FRG; E-mail: Mewes@mips.embnet.org
FEATURES             Location/Qualifiers
     source          1. .6323
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4932"
                     /chromosome="VII"
     gene            join(2261. .2263,2904. .5480)
                     /gene="MPT5"
     CDS             join(2261. .2263,2904. .5480)
                     /gene="MPT5"
                     /note="ORF YGL178w"
                     /codon_start=1
                     /protein_id="CAA96889.1"
                     /db_xref="GI:1322786"
                     /db_xref="GOA:P39016"
                     /db_xref="SGD:S000003146"
                     /db_xref="UniProt/Swiss-Prot:P39016"
                     /translation="MINNEPFPSADSASILTTSTSNNSLMSYNHQPQLSINSVQSLLE
                     PVTPPPLGQMNNKRNHQKAHSLDLSGFNQFISSTQSPLALMNNTSTSNSANSFSPNPN
                     AASNSTGLSASMANPPAILPLINEFDLEMDGPRRKSSHDFTVVAPSNSGVNTSSLIME
                     TPSSSVTPAASLRNFSNSNNAASKCGVDNSSFGLSSSTSSSMVEISALPLRDLDYIKL
                     ATDQFGCRFLQKKLETPSESNMVRDLMYEQIKPFFLDLILDPFGNYLVQKLCDYLTAE
                     QKTLLIQTIYPNVFQISINQYGTRSLQKIIDTVDNEVQIDLIIKGFSQEFTSIEQVVT
                     LINDLNGNHVIQKCIFKFSPSKFGFIIDAIVEQNNIITISTHKHGCCVLQKLLSVCTL
                     QQIFKISVKIVQFLPGLINDQFGNYIIQFLLDIKELDFYLLAELFNRLSNELCQLSCL
                     KFSSNVVEKFIKKLFRIITGFIVNNNGGASQRTAVASDDVINASMNILLTTIDIFTVN
                     LNVLIRDNFGNYALQTLLDVKNYSPLLAYNKNSNAIGQNSSSTLNYGNFCNDFSLKIG
                     NLIVLTKELLPSIKTTSYAKKIKLKVKAYAEATGIPFTDISPQVTAMSHNNLQTINNE
                     NKNPHNKNSHNHNHNHNHNHAHNNNNNNNQKSHTRHFSLPANAYHRRSNSSVTNNFSN
                     QYAQDQKIHSPQQIMNFNQNAYPSMGAPSFNSQTNPPLVSHNSLQNFDNRQFANLMAH
                     PNSAAPIHSFSSSNITNVNPNVSRGFKQPGFMMNETDKINANHFSPYSNANSQNFNES
                     FVPRMQYQTEGANWDSSLSMKSQHIGQGPYNQVNMSRNASISNMPAMNTARTSDELQF
                     TLP"
     exon            <2261. .2263
                     /gene="MPT5"
                     /number=1
     intron          2264. .2903
                     /gene="MPT5"
                     /number=1
     CDS             2468. .2815
                     /gene="MPT5"
                     /note="unnamed protein product; ORF YGL177w"
                     /codon_start=1
                     /protein_id="CAA96890.1"
                     /db_xref="GI:1322787"
                     /db_xref="SGD:S000003145"
                     /db_xref="UniProt/Swiss-Prot:P53105"
                     /translation="MPIRIFSKEGLRERGASGKNEQKKKKKEKIRRSMKNLCRFVESY
                     KDFTLYGYRFHSKIKHKFCVLSSSFPVLFFCHPLSPLKSNFIDRFTLDSHASQATWTL
                     YMVVIVLFLSYRN"
     exon            2904. .>5480
                     /gene="MPT5"
                     /number=2
ORIGIN

Alignment Scores:
Pred. No.:                1.38e+03        Length:         6323
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
```

```
DB:                    15             Gaps:          0

US-09-858-332G-28 (1-6) x SCYGL178W (1-6323)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       4800 CATAATCATAATCATAAT 4817

RESULT 348
AX594150
LOCUS       AX594150               6525 bp    DNA     linear   PAT 13-FEB-2003
DEFINITION  Sequence 3 from Patent WO02083910.
ACCESSION   AX594150
VERSION     AX594150.1  GI:28375380
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Farmer,A.A.
  TITLE     Sequence specific recombinase-based methods for producing intron
            containing vectors and compositions for use in practicing the same
  JOURNAL   Patent: WO 02083910-A 3 24-OCT-2002;
            Clontech Laboratories Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .6525
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Synthetic Vector"
ORIGIN

Alignment Scores:
Pred. No.:             1.42e+03       Length:        6525
Score:                 42.00          Matches:       6
Percent Similarity:    100.00%        Conservative:  0
Best Local Similarity: 100.00%        Mismatches:    0
Query Match:           100.00%        Indels:        0
DB:                    6              Gaps:          0

US-09-858-332G-28 (1-6) x AX594150 (1-6525)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1728 CATAATCATAATCATAAT 1745

RESULT 349
MIGCYP1G/c
LOCUS       MIGCYP1G               6837 bp    DNA     linear   VRT 16-MAY-1996
DEFINITION  Microgadus tomcod cytochrome P4501A1 (CYP1A1) gene, complete cds.
ACCESSION   L41917
VERSION     L41917.1  GI:797269
KEYWORDS    cytochrome P4501A1.
SOURCE      Microgadus tomcod (Atlantic tomcod)
  ORGANISM  Microgadus tomcod
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae; Microgadus.
REFERENCE   1  (bases 1 to 6837)
  AUTHORS   Roy,N.K., Kreamer,G.L., Konkle,B., Grunwald,C. and Wirgin,I.
  TITLE     Characterization and prevalence of a polymorphism in the 3'
            untranslated region of cytochrome P4501A1 in cancer-prone Atlantic
            tomcod
  JOURNAL   Arch. Biochem. Biophys. 322 (1), 204-213 (1995)
   PUBMED   7574676
COMMENT     On May 5, 1995 this sequence version replaced gi:793750.
FEATURES             Location/Qualifiers
     source          1. .6837
                     /organism="Microgadus tomcod"
                     /mol_type="genomic DNA"
                     /strain="Hudson River Population"
                     /db_xref="taxon:34823"
                     /clone="11A32"
                     /tissue_type="liver"
                     /dev_stage="adult"
                     /note="(vector AT/lambda Dash II)"
     misc_feature    73. .77
                     /note="dioxin responsive element core sequence"
     misc_feature    166. .170
                     /note="dioxin responsive element core sequence"
     misc_feature    556. .562
                     /note="glucocorticoid responsive element"
     misc_feature    600. .606
                     /note="glucocorticoid responsive element"
     misc_feature    1692. .1696
                     /note="dioxin responsive element core sequence"
     misc_feature    1770. .1774
                     /note="dioxin responsive element core sequence"
     misc_feature    2285. .2289
                     /note="dioxin responsive element core sequence"
     TATA_signal     2500. .2505
                     /gene="CYP1A1"
     prim_transcript 2530. .>6834
                     /gene="CYP1A1"
     gene            join(2543. .2628,3094. .3942,4215. .4341,4451. .4540,
                     4955. .5078,5431. .5517,5704. .6837)
                     /gene="CYP1A1"
     mRNA            join(2543. .2628,3094. .3942,4215. .4341,4451. .4540,
                     4955. .5078,5431. .5517,5704. .6837)
                     /gene="CYP1A1"
     exon            2543. .2628
                     /gene="CYP1A1"
                     /number=1
     intron          2629. .3093
                     /gene="CYP1A1"
                     /number=1
     exon            3094. .3942
                     /gene="CYP1A1"
                     /number=2
     CDS             join(3115. .3942,4215. .4341,4451. .4540,4955. .5078,
                     5431. .5517,5704. .5995)
                     /gene="CYP1A1"
                     /codon_start=1
                     /product="cytochrome P450 1A1"
                     /protein_id="AAB00085.1"
                     /db_xref="GI:797270"
                     /translation="MALMILPLIGSVSVSETLVAMITVCMIYMLMKFLHPDVPEAPPA
                     PGPQALPIIGNVWSWENGLPEPHGHGQRYGDILQIQIGMRSVVVLSGHETVRQALIKQ
                     GHDLRPPDLYSFQFINDGKSLAFSTDQAGVWSPPKAGHECPALLFHARGHHAAVLLHV
                     GGACLQGGRLPRQAAVQRHGTDASFDPFRHIVVSVANVICGMCFGRRYGHEDQELLSL
                     VNLTDEFGKVVGSGNLADFIPLLRFLPNATMKRFMAINERFMTFVQKIVTEHYNTYDK
                     DNIRDITDSLIDHCEDRKLDENSNIQMSDEKIVGIVNDLFGAGFDTVSTALSWSVMYL
                     VAHPEIQERLHQEIKDKVGLSRSPVLTDRHNLPILEAFIFEIFRHSSFLPFTIPHCAT
                     KDTSLDGYFIPKDTCVFINQWQINHDPELWKEPSTFNPDRFLSADASELNKLAGEKVM
                     LFGMGKRRCIGEMVARNEVFLFLAILVQRLTFHAVPGEPLDMTPEYGLTMKHKRCHLR
                     ATVRTTE"
     intron          3943. .4214
                     /gene="CYP1A1"
                     /number=2
     exon            4215. .4341
                     /gene="CYP1A1"
                     /number=3
     intron          4342. .4450
                     /gene="CYP1A1"
                     /number=3
     exon            4451. .4540
                     /gene="CYP1A1"
                     /number=4
     intron          4541. .4954
                     /gene="CYP1A1"
                     /number=4
     exon            4955. .5078
                     /gene="CYP1A1"
                     /number=1
     intron          5079. .5430
                     /gene="CYP1A1"
```

```
                     /number=5
     exon            5431. .5517
                     /gene="CYP1A1"
                     /number=6
     intron          5518. .5703
                     /gene="CYP1A1"
                     /number=6
     exon            5704. .6837
                     /gene="CYP1A1"
                     /number=7
     polyA_signal    6802. .6807
                     /gene="CYP1A1"
     polyA_signal    6812. .6819
                     /gene="CYP1A1"
ORIGIN

Alignment Scores:
Pred. No.:              1.49e+03        Length:         6837
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     5               Gaps:           0

US-09-858-332G-28 (1-6) x MIGCYP1G (1-6837)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1171 CATAATCACAATCACAAT 1154

RESULT 350
AF433042
LOCUS       AF433042               6875 bp    DNA     circular SYN 11-NOV-2001
DEFINITION  Cloning vector pWS31, complete sequence.
ACCESSION   AF433042
VERSION     AF433042.1  GI:16903089
KEYWORDS    .
SOURCE      Cloning vector pWS31
  ORGANISM  Cloning vector pWS31
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 6875)
  AUTHORS   Sundaresan,V., Springer,P., Volpe,T., Haward,S., Jones,J.D.,
            Dean,C., Ma,H. and Martienssen,R.
  TITLE     Patterns of gene action in plant development revealed by enhancer
            trap and gene trap transposable elements
  JOURNAL   Genes Dev. 9 (14), 1797-1810 (1995)
   PUBMED   7622040
REFERENCE   2  (bases 1 to 6875)
  AUTHORS   de la Bastide,M., Preston,R., O'Shaughnessy,A., May,B.,
            Sundaresan,V., Martienssen,R. and McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-OCT-2001) Cold Spring Harbor Laboratory, 1 Bungtown
            Rd, Cold Spring Harbor, NY 11724, USA
FEATURES             Location/Qualifiers
     source          1. .6875
                     /organism="Cloning vector pWS31"
                     /mol_type="other DNA"
                     /db_xref="taxon:176659"
     source          64. .1852
                     /organism="Cloning vector pWS31"
                     /mol_type="other DNA"
                     /db_xref="taxon:176659"
                     /transposon="Ds transposable element"
                     /note="5' fragment"
     source          6458. .6649
                     /organism="Cloning vector pWS31"
                     /mol_type="other DNA"
                     /db_xref="taxon:176659"
                     /transposon="Ds transposable element"
                     /note="3' fragment; transferred to T-DNA vector pAJ8 for
                     transformation of Arabidopsis thaliana"
     source          6670. .6861
                     /organism="Cloning vector pWS31"
                     /mol_type="other DNA"
                     /db_xref="taxon:176659"
                     /transposon="Ds transposable element"
                     /note="3' fragment"
     gene            2376. .3170
                     /gene="npt2"
     CDS             2376. .3170
                     /gene="npt2"
                     /codon_start=1
                     /transl_table=11
                     /product="neomycin phosphotransferase"
                     /protein_id="AAL30400.1"
                     /db_xref="GI:16903090"
                     /translation="MIEQDGLHAGSPAAWVERLFGYDWAQQTIGCSDAAVFRLSAQGR
                     PVLFVKTDLSGALNELQDEAARLSWLATTGVPCAAVLDVVTEAGRDWLLLGEVPGQDL
                     LSSHLAPAEKVSIRLNAMRRLHTLDPATCPFDHQAKHRIERARTRMEAGLVDQDDLDE
                     EHQGLAPAELFARLKARMPDGEDLVVTHGDACLPNIMVENGRFSGFIDCGRLGVADRY
                     QDIALATRDIAEELGGEWADRFLVLYGIAAPDSQRIAFYRLLDEFF"
     gene            complement(4550. .6361)
                     /gene="uidA"
     CDS             complement(4550. .6361)
                     /gene="uidA"
                     /codon_start=1
                     /transl_table=11
                     /product="beta-glucuronidase"
                     /protein_id="AAL30401.1"
                     /db_xref="GI:16903091"
                     /translation="MVRPVETPTREIKKLDGLWAFSLDRENCGIDQRWWESALQESRA
                     IAVPGSFNDQFADADIRNYAGNVWYQREVFIPKGWAGQRIVLRFDAVTHYGKVWVNNQ
                     EVMEHQGGYTPFEADVTPYVIAGKSVRITVCVNNELNWQTIPPGMVITDENGKKKQSY
                     FHDFFNYAGIHRSVMLYTTPNTWVDDITVVTHVAQDCNHASVDWQVVANGDVSVELRD
                     ADQQVVATGQGTSGTLQVVNPHLWQPGEGYLYELCVTAKSQTECDIYPLRVGIRSVAV
                     KGEQFLINHKPFYFTGFGRHEDADLRGKGFDNVLMVHDHALMDWIGANSYRTSHYPYA
                     EEMLDWADEHGIVVIDETAAVGFNLSLGIGFEAGNKPKELYSEEAVNGETQQAHLQAI
                     KELIARDKNHPSVVMWSIANEPDTRPQGAREYFAPLAEATRKLDPTRPITCVNVMFCD
                     AHTDTISDLFDVLCLNRYYGWYVQSGDLETAEKVLEKELLAWQEKLHQPIIITEYGVD
                     TLAGLHSMYTDMWSEEYQCAWLDMYHRVFDRVSAVVGEQVWNFADFATSQGILRVGGN
                     KKGIFTRDRKPKSAAFLLQKRWTGMNFGEKPQQGGKQ"
     misc_feature    complement(6381. .6414)
                     /note="CaMV 35S promoter -46 region"
ORIGIN

Alignment Scores:
Pred. No.:              1.5e+03         Length:         6875
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     11              Gaps:           0

US-09-858-332G-28 (1-6) x AF433042 (1-6875)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1483 CACAACCACAACCACAAC 1500

RESULT 351
AF433043
LOCUS       AF433043               6928 bp    DNA     circular SYN 11-NOV-2001
DEFINITION  Cloning vector pWS32, complete sequence.
ACCESSION   AF433043
VERSION     AF433043.1  GI:16903092
KEYWORDS    .
SOURCE      Cloning vector pWS32
  ORGANISM  Cloning vector pWS32
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 6928)
  AUTHORS   Sundaresan,V., Springer,P., Volpe,T., Haward,S., Jones,J.D.,
            Dean,C., Ma,H. and Martienssen,R.
  TITLE     Patterns of gene action in plant development revealed by enhancer
            trap and gene trap transposable elements
  JOURNAL   Genes Dev. 9 (14), 1797-1810 (1995)
   PUBMED   7622040
```

```
REFERENCE   2  (bases 1 to 6928)
  AUTHORS   de la Bastide,M., Preston,R., O'Shaughnessy,A., May,B.,
            Sundaresan,V., Martienssen,R. and McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-OCT-2001) Cold Spring Harbor Laboratory, 1 Bungtown
            Rd, Cold Spring Harbor, NY 11724, USA
FEATURES             Location/Qualifiers
     source          1. .6928
                     /organism="Cloning vector pWS32"
                     /mol_type="other DNA"
                     /db_xref="taxon:176660"
     source          64. .1852
                     /organism="Cloning vector pWS32"
                     /mol_type="other DNA"
                     /db_xref="taxon:176660"
                     /transposon="Ds transposable element"
                     /note="5' fragment"
     source          6511. .6702
                     /organism="Cloning vector pWS32"
                     /mol_type="other DNA"
                     /db_xref="taxon:176660"
                     /transposon="Ds transposable element"
                     /note="3' fragment; transferred to T-DNA vector pAJ8 for
                     transformation of Arabidopsis thaliana"
     source          6720. .6911
                     /organism="Cloning vector pWS32"
                     /mol_type="other DNA"
                     /db_xref="taxon:176660"
                     /transposon="Ds transposable element"
                     /note="3' fragment"
     gene            2376. .3170
                     /gene="npt2"
     CDS             2376. .3170
                     /gene="npt2"
                     /codon_start=1
                     /transl_table=11
                     /product="neomycin phosphotransferase"
                     /protein_id="AAL30402.1"
                     /db_xref="GI:16903093"
                     /translation="MIEQDGLHAGSPAAWVERLFGYDWAQQTIGCSDAAVFRLSAQGR
                     PVLFVKTDLSGALNELQDEAARLSWLATTGVPCAAVLDVVTEAGRDWLLLGEVPGQDL
                     LSSHLAPAEKVSIMADAMRRLHTLDPATCPFDHQAKHRIERARTRMEAGLVDQDDLDE
                     EHQGLAPAELFARLKARMPDGEDLVVTHGDACLPNIMVENGRFSGFIDCGRLGVADRY
                     QDIALATRDIAEELGGEWADRFLVLYGIAAPDSQRIAFYRLLDEFF"
     gene            complement(4550. .6361)
                     /gene="uidA"
     CDS             complement(4550. .6361)
                     /gene="uidA"
                     /codon_start=1
                     /transl_table=11
                     /product="beta-glucuronidase"
                     /protein_id="AAL30403.1"
                     /db_xref="GI:16903094"
                     /translation="MVRPVETPTREIKKLDGLWAFSLDRENCGIDQRWWESALQESRA
                     IAVPGSFNDQFADADIRNYAGNVWYQREVFIPKGWAGQRIVLRFDAVTHYGKVWVNNQ
                     EVMEHQGGYTPFEADVTPYVIAGKSVRITVCVNNELNWQTIPPGMVITDENGKKKQSY
                     FHDFFNYAGIHRSVMLYTTPNTWVDDITVVTHVAQDCNHASVDWQVVANGDVSVELRD
                     ADQQVVATGQGTSGTLQVVNPHLWQPGEGYLYELCVTAKSQTECDIYPLRVGIRSVAV
                     KGEQFLINHKPFYFTGFGRHEDADLRGKGFDNVLMVHDHALMDWIGANSYRTSHYPYA
                     EEMLDWADEHGIVVIDETAAVGFNLSLGIGFEAGNKPKELYSEEAVNGETQQAHLQAI
                     KELIARDKNHPSVVMWSIANEPDTRPQGAREYFAPLAEATRKLDPTRPITCVNVMFCD
                     AHTDTISDLFDVLCLNRYYGWYVQSGDLETAEKVLEKELLAWQEKLHQPIIITEYGVD
                     TLAGLHSMYTDMWSEEYQCAWLDMYHRVFDRVSAVVGEQVWNFADFATSQGILRVGGN
                     KKGIFTRDRKPKSAAFLLQKRWTGMNFGEKPQQGGKQ"
     intron          complement(6364. .6615)
                     /note="synthetic GPA1 intron containing splice donors and
                     acceptors in 3 reading frames"
ORIGIN

Alignment Scores:
Pred. No.:              1.51e+03        Length:         6928
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     11              Gaps:           0

US-09-858-332G-28 (1-6) x AF433043 (1-6928)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1483 CACAACCACAACCACAAC 1500

RESULT 352
ZMU11059/c
LOCUS       ZMU11059                7313 bp    DNA     linear   PLN 17-JUN-2005
DEFINITION  Zea mays A188 retrotransposon gag gene, complete cds.
ACCESSION   U11059
VERSION     U11059.1  GI:507844
KEYWORDS    .
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 7313)
  AUTHORS   Hu,W., Das,O.P. and Messing,J.
  TITLE     Zeon-1, a member of a new maize retrotransposon family
  JOURNAL   Mol. Gen. Genet. 248 (4), 471-480 (1995)
   PUBMED   7565611
REFERENCE   2  (bases 1 to 7313)
  AUTHORS   Messing,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUN-1994) Joachim Messing, Waksman Institute, Rutgers
            State University, Hoes Lane, Piscataway, NJ 08855, USA
FEATURES             Location/Qualifiers
     source          1. .7313
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="A188"
                     /db_xref="taxon:4577"
                     /chromosome="7"
                     /tissue_type="leaf"
     repeat_region   1. .7313
                     /transposon="retrotransposon"
     LTR             1. .649
     TATA_signal     304. .310
     misc_feature    651. .659
                     /note="PBS site"
     gene            921. .2048
                     /gene="gag"
     CDS             921. .2048
                     /gene="gag"
                     /codon_start=1
                     /product="gag"
                     /protein_id="AAA93147.1"
                     /db_xref="GI:507845"
                     /translation="MARLADLQRQIDEASEEVRHLVQDDQNRRPPRRELHQEGFYNED
                     DWYEDFHHGNFAFDDASPLSTELQATPWPQSYKPPQLPMYDGHTDPKQFLMSYEATIS
                     SYGGNTAVMAKSFVMAVKSVAQTWYSSLRPGTITSWQKLKDMLITSFQGFQTKPVTAQ
                     ALFQCTQDHEEYLQAYVRRFLRLRAQAPTVPNEIVIEAMIKGLRSGHSAQYFARKPPQ
                     TLEKLLQKMDEYIRADNDFRQRREEAFRFSEMTRGFGGRFHTRHVRSIHNSTQSDDRG
                     SQQQRPQYSLQASGQQQSPVRPPAPRGRGARGFGGRFGDQPRRIYCIFCGEDKGHTTR
                     MCHVTIQKQKEIAEAAASVASASQPQASWQQPPPPPPTQRG"
     misc_feature    6641. .6650
                     /note="PPTsite"
     LTR             6651. .7313
     TATA_signal     6969. .6975
ORIGIN

Alignment Scores:
Pred. No.:              1.59e+03        Length:         7313
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
```

```
DB:                    15            Gaps:          0

US-09-858-332G-28 (1-6) x ZMU11059 (1-7313)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        508 CATAATCATAATCATAAT 491

RESULT 353
AR097039
LOCUS       AR097039                7334 bp    DNA     linear   PAT 14-FEB-2001
DEFINITION  Sequence 1 from patent US 6071518.
ACCESSION   AR097039
VERSION     AR097039.1  GI:12805769
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7334)
  AUTHORS   Petersen,C.
  TITLE     GP900 glycoprotein and fragments for treatment and
            detection/diagnosis of cryptosporidium
  JOURNAL   Patent: US 6071518-A 1 06-JUN-2000;
FEATURES             Location/Qualifiers
     source          1. .7334
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

Alignment Scores:
Pred. No.:              1.59e+03        Length:         7334
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x AR097039 (1-7334)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       2910 CACAACCACAACCACAAC 2927

RESULT 354
AR279258
LOCUS       AR279258                7334 bp    DNA     linear   PAT 10-APR-2003
DEFINITION  Sequence 1 from patent US 6514697.
ACCESSION   AR279258
VERSION     AR279258.1  GI:29713986
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7334)
  AUTHORS   Petersen,C., Barnes,D.A., Nelson,R.C. and Gut,J.
  TITLE     Methods for detection of Crytosporidium species and isolates and
            for diagnosis of Cryptosporidium infections
  JOURNAL   Patent: US 6514697-A 1 04-FEB-2003;
            The Regents of the University of California; Oakland, CA
FEATURES             Location/Qualifiers
     source          1. .7334
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              1.59e+03        Length:         7334
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x AR279258 (1-7334)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       2910 CACAACCACAACCACAAC 2927

RESULT 355
AX594151
LOCUS       AX594151                7487 bp    DNA     linear   PAT 13-FEB-2003
DEFINITION  Sequence 4 from Patent WO02083910.
ACCESSION   AX594151
VERSION     AX594151.1  GI:28375381
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Farmer,A.A.
  TITLE     Sequence specific recombinase-based methods for producing intron
            containing vectors and compositions for use in practicing the same
  JOURNAL   Patent: WO 02083910-A 4 24-OCT-2002;
            Clontech Laboratories Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .7487
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="Synthetic Vector"
ORIGIN

Alignment Scores:
Pred. No.:              1.63e+03        Length:         7487
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x AX594151 (1-7487)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       2316 CATAATCATAATCATAAT 2333

RESULT 356
AB055064
LOCUS       AB055064                7635 bp    DNA     linear   SYN 21-MAY-2003
DEFINITION  Synthetic transposable element dAc-I-RS DNA, complete sequence.
ACCESSION   AB055064
VERSION     AB055064.1  GI:13359266
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Machida,C., Onouchi,H., Koizumi,J., Hamada,S., Semiarti,E.,
            Torikai,S. and Machida,Y.
  TITLE     Characterization of the transposition pattern of the Ac element in
            Arabidopsis thaliana using endonuclease I-SceI
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 94 (16), 8675-8680 (1997)
   PUBMED   11038561
REFERENCE   2
  AUTHORS   Semiarti,E., Onouchi,H., Torikai,S., Ishikawa,T., Machida,Y. and
            Machida,C.
  TITLE     The transposition pattern of the Ac element in tobacco cultured
            cells
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 7635)
  AUTHORS   Machida,C. and Machida,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JAN-2001) Yasunori Machida, Nagoya University,
            Division of Biological Science, Graduate School of Science;
            Chikusa-ku, Furo-cho, Nagoya, Aichi 464-8602, Japan
```

```
          (E-mail:yas@bio.nagoya-u.ac.jp, Tel:81-52-789-5040,
          Fax:81-52-789-2966)
FEATURES             Location/Qualifiers
     source          1. .7635
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="non-autonomous derivative of Ac
                     transposable element dAc-I-RS"
     misc_feature    1. .1789
                     /note="sequence of the Ac element in maize, containing the
                     terminal inverted repeat"
     misc_feature    1807. .1864
                     /note="RS recombination site recognized by R protein from
                     Zygosaccharomyces rouxii"
     promoter        2016. .2380
                     /note="the 35S promoter from cauliflower mosaic virus"
     gene            2390. .3415
                     /gene="HPT"
     CDS             2390. .3415
                     /gene="HPT"
                     /codon_start=1
                     /transl_table=11
                     /product="hygromycin phosphotransferase"
                     /protein_id="BAB33367.1"
                     /db_xref="GI:13359267"
                     /translation="MKKPELTATSVEKFLIEKFDSVSDLMQLSEGEESRAFSFDVGGR
                     GYVLRVNSCADGFYKDRYVYRHFASAALPIPEVLDIGEFSESLTYCISRRAQGVTLQD
                     LPETELPAVLQPVAEAMDAIAAADLSQTSGFGPFGPQQGIGQYTTWRDFICAIADPHVY
                     HWQTVMDDTVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTDNGRITAVIDWS
                     EAMFGDSQYEVANIFFWRPWLACMEQQTRYFERRHPELAGSPRLRAYMLRIGLDQLYQ
                     SLVDGNFDDAAWAQGRCDAIVRSGAGTVGRTQIARRSAAVWTDGCVEVLADSGNRRPS
                     TRPRAKE"
     terminator      3436. .4560
                     /note="the terminator of the gene for nopaline synthase"
     misc_feature    4575. .4632
                     /note="RS recombination site recognized by R protein from
                     Zygosaccharomyces rouxii"
     misc_feature    4683. .4700
                     /note="cleavage site for endonuclease I-SceI"
     misc_feature    5149. .5172
                     /note="cos sequence"
     misc_feature    6496. .7635
                     /note="sequence of the Ac element in maize, containing the
                     terminal inverted repeat"
ORIGIN

Alignment Scores:
Pred. No.:              1.66e+03        Length:         7635
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     11              Gaps:           0

US-09-858-332G-28 (1-6) x AB055064 (1-7635)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1420 CACAACCACAACCACAAC 1437

RESULT 357
AY160101
LOCUS       AY160101               7801 bp    DNA     linear   INV 23-OCT-2002
DEFINITION  Dictyostelium discoideum nucleotide exchange factor RasGEF L (gefL)
            gene, complete cds.
ACCESSION   AY160101
VERSION     AY160101.1  GI:24286698
KEYWORDS    .
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 7801)
  AUTHORS   Wilkins,A., Szafranski,K., Gloeckner,G., Harrisingh,M.,
            Deenadayalan,B., Mueller,R., Eichinger,L., Noegel,A.A. and
            Insäll,R.
  TITLE     The family of rasGEF genes in Dictyostelium discoideum
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 7801)
  AUTHORS   Szafranski,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-2002) Abt. Genomanalyse, Institut fuer Molekulare
            Biotechnologie, Beutenbergstrasse 11, D-07745 Jena, Germany
FEATURES             Location/Qualifiers
     source          1. .7801
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
     gene            <296. .>5023
                     /gene="gefL"
     mRNA            <296. .>5023
                     /gene="gefL"
                     /product="nucleotide exchange factor RasGEF L"
     CDS             296. .5023
                     /gene="gefL"
                     /note="GefL"
                     /codon_start=1
                     /product="nucleotide exchange factor RasGEF L"
                     /protein_id="AAN46881.1"
                     /db_xref="GI:24286699"
                     /translation="MGSPIYKSFQSLYETLLDDLDYSKKRREELLNLSVEEKWIFLQR
                     IKPDCTNSLVSNGVYQLQHQISSNDLKTLISSPPKLNSSQQLLSSSSSSSSLSSSSTS
                     ATHSLQNKSLPSISVCPKSPRTHMKVLSASSLFTLEDITKLRQQLVSKPVSWVINFAE
                     STGIVSILSMITTYSQNLINNVNKDQPQSQPQPQPQPQPRPQPQQPQPQPQQPQPQQLLP
                     LLSQLQPQQQQQQQQQTQQPQPQLQAQTKSDDHFIINECLGCIKLLVELEMQSLLLND
                     TAEIIFPLILSSQVSTKRMTFEVLELICKIFSLGSTIVLESIKNYSKKINEPLTKVFS
                     HITSPLDSEYPADLKVSCMSLINHLIGSNHLSLEARIETRSKFLSNDILLIIKNLRID
                     GEGIETSSWKQLELEMKQFESMMVQDDNQTSNAEANELVSNELSQSIDNNNNNNNNNN
                     NNSSTNLVNNNNNNNINSNCNSIINGGNSSDSLINNSVGTGVGSIMGVGSFMGVNEIE
                     YDSPRKEFDELKSSNNNNSNSNSVNNFINNITEPIMPLTPTEGKLVVIVSMQQTRQQA
                     QLQQQLQQQLEEEDDDEVPLSSSVATTASTTANTTLIEPNHLPIQTREVIVSVQRKSQ
                     AGDVIRSLLASNSNLKELGEWGLFVNDSNSNNNNNNNNNNNNKLSTNVIGRFLKDDEY
                     IFDQMTTSSTSGSLSSDDGATPRLMTRSIPRFKEYRLKMIPWKVRVSLEKVKEMGLVL
                     NANNTANGQFNESKFIAEAPMDPTMTCTGLVYHLIKKYLPNVNDRYCLEADDFGLYLD
                     TNAGGLMWGGGTGGGYWLEPLEKLHDYEIFKDPKSTVALKLRPKGVKLKFADGTYEIF
                     RLDLMSPTEKIFQEISEKVIDSNNNINLSYYGIFVDRQQGGANGNGISDSPILVSQSS
                     STSLIEESKKQRKPEWLEKGLPLYHYKINSRCCLRFALRPSNITLLIDISLFKSFQQQ
                     QQQQQQQQQQSTSSSSSSSSSSVSSNNNSTNNSNNLIGSLNNSINTISELDESIDDN
                     GLSSSSSDKEDQISTPSTSNLLSKFQQVDESISNLKFQLPQPIQSKQIIVQLPLHLPL
                     GEALETDENWIETGIEREKCSFHFDSFDGILMNQTQPLNTQNFTEKNRICVLMADELE
                     INSEQTNHNKTNIWDEPFDSSTIICDPDNPGQIRAATLNKLLEQATNNIEKDRELMNI
                     LLMTYMSFATSDILLDKLVERYNVPEHEKEQKSVVQLHTIVFLKNWLEQQAPQAASGG
                     GLEEKFLVRLSEFVDKQISLDGYTKIVPQLKKWIDSALKEKRAYAMPEAARPNQLTTK
                     LPSLTSSLLDDELFVAQQLTLREYETFKRIQAVEFLGQAWGKPKLQYKAPNLLKMIER
                     FNKISTGVSTAILSQPKIKTRVKLICRFIKIAQHCRDLNNYHLLTAFMAGIRNSNVVR
                     LRLSWAKVPKKHLHILEDLEKIMSMEGSFKEFRTKMAETIPPCIPYLGVYLKDLTFIE
                     EGNPDSINSLINWGKKKLIYNIISIIQRCQQVSYDFGNQNQSQVAQKSDLVLNYFDSL
                     PIALDELLYKISLNLEPRSNFNN"
ORIGIN

Alignment Scores:
Pred. No.:              1.69e+03        Length:         7801
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-858-332G-28 (1-6) x AY160101 (1-7801)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        855 CACAATCACAACCACAAC 872

RESULT 358
```

```
AC014533
LOCUS       AC014533               7916 bp    DNA     linear   HTG 16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC014533
VERSION     AC014533.1  GI:6436802
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 7916)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10212889 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .7916
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
ORIGIN

Alignment Scores:
Pred. No.:                1.72e+03        Length:          7916
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       14              Gaps:            0

US-09-858-332G-28 (1-6) x AC014533 (1-7916)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        660 CACAACCACAACCACAAT 677

RESULT 359
SCDNAGMH
LOCUS       SCDNAGMH               7951 bp    DNA     linear   PLN 18-APR-2005
DEFINITION  S.cerevisiae G1618 gene and MPT5/HTR1 gene.
ACCESSION   X83690
VERSION     X83690.1  GI:794143
KEYWORDS    G1618 gene; HTR1 gene; MPT5 gene; mutation suppressor; ser/thr
            protein kinase; upstream activation sequence.
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1
  AUTHORS   Coglievina,M., Bertani,I., Klima,R., Zaccaria,P. and Bruschi,C.V.
  TITLE     The DNA sequence of a 7941 bp fragment of the left arm of
            chromosome VII of Saccharomyces cerevisiae contains four open
            reading frames including the multicopy suppressor gene of the pop2
            mutation and a putative serine/threonine protein kinase gene
  JOURNAL   Yeast 11 (8), 767-774 (1995)
   PUBMED   7668046
REFERENCE   2  (bases 1 to 7951)
  AUTHORS   Bruschi,C.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-1995) C.V. Bruschi, Microbiology, International
            Centre for Genetic Engineering and Biotechnology, AREA Science
            Park, Padriciano 99, I-34012 Trieste, ITALY
COMMENT     Sequence overlapping with those under the accession numbers D26184
            and D25541 and also with est 102637 of the TIGR database.
FEATURES             Location/Qualifiers
     source          1. .7951
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /strain="FY1679"
                     /db_xref="taxon:4932"
                     /chromosome="VII"
                     /map="left arm"
     gene            complement(87. .1769)
                     /gene="biE560"
     CDS             complement(87. .1769)
                     /gene="biE560"
                     /codon_start=1
                     /product="putative ser /thr protein kinases"
                     /protein_id="CAA58659.1"
                     /db_xref="GI:794144"
                     /db_xref="GOA:P43637"
                     /db_xref="InterPro:IPR000719"
                     /db_xref="InterPro:IPR002290"
                     /db_xref="InterPro:IPR008271"
                     /db_xref="InterPro:IPR011009"
                     /db_xref="SGD:S000003147"
                     /db_xref="UniProt/Swiss-Prot:P43637"
                     /translation="MVLLKEPVQPLPRSSLLYNNASNSSSRIKETRKVKLLYNPLTKR
                     QILNNFEILATLGNGQYGKVKLARDLGTGALVAIKILNRFEKRSGYSLQLKVENPRVN
                     QEIEVMKRCHHENVVELYEILNDPESTKVYLVLEYCSRGPVKWCPENKMEIKAVGPSI
                     LTFQQSRKVVLDVVSGLEYLHSQGITHRDIKPSNLLISSNGTVKISDFGVAMSTATGS
                     TNIQSSHEQLLKSRALGTPAFFAPELCSTEKEYSCSSAIDIWSLGVTIYCLLFGKLPF
                     NANSGLELFDSIINKPLEFPSYEEMLNGATSGITMEEYTDAKDLLKKLLQKDPDKRIK
                     LADIKVHPFMCHYGKSDAASVLTNLETFHELKVSPPSSCKRVELVSLPVNSSFASLDS
                     VYMENFDHNNLRTGADRNSTYSPSIYDANTLSPSAYHNIGSRESSYSSFSSFTSSTAF
                     ASQISIQDAPAIGDQQCLIGESGSSLRVNSCEFPQYTTMSPVGEYPFESTEASLSSTL
                     TPVGNVPQRIKAHLVEGKSNSKDDLRIEADASLVFEASDAQRTRRRMSLYKL"
     gene            1000. .1377
                     /gene="biB125"
     CDS             1000. .1377
                     /gene="biB125"
                     /note="putative protein"
                     /codon_start=1
                     /protein_id="CAA58661.1"
                     /db_xref="GI:1209469"
                     /db_xref="UniProt/TrEMBL:Q05671"
                     /translation="MADEQEYSFSVEQSSGAKKAGVPKALDFKSCSCDDCILVLPVAV
                     DIATPKSDIFTVPFDEINRFDGLISRCVIPWECKYSRPDTTSNTTFLDCWKVSIDGPT
                     AFISILFSGHHLTGPLEQYSNTR"
     gene            1881. .1888
                     /gene="MPT5=HTR1"
     enhancer        1881. .1888
                     /gene="MPT5=HTR1"
                     /note="HAP2/3/4 putative UAS"
     gene            join(4030. .4032,4673. .7073,7912. .7946)
                     /gene="MPT5=HTR1"
     CDS             join(4030. .4032,4673. .7073,7912. .7946)
                     /gene="MPT5=HTR1"
                     /function="multicopy suppressor of pop2 mutation"
                     /codon_start=1
                     /protein_id="CAA58660.1"
                     /db_xref="GI:794145"
                     /db_xref="GOA:P39016"
                     /db_xref="UniProt/Swiss-Prot:P39016"
                     /translation="MINNEPFPSADSASILTTSTSNNSLMSYNHQPQLSINSVQSLLE
                     PVTPPPLGQMNNKRNHQKAHSLDLSGFNQFISSTQSPLALMNNTSTSNSANSFSPNPN
                     AASNSTGLSASMANPPAILPLINEFDLEMDGPRRKSSHDFTVVAPSNSGVNTSSLIME
                     TPSSSVTPAASLRNFSNSNNAASKCGVDNSSFGLSSSTSSSMVEISALPLRDLDYIKL
                     ATDQFGCRFLQKKLETPSESNMVRDLMYEQIKPFFLDLILDPFGNYLVQKLCDYLTAE
                     QKTLLIQTIYPNVFQISINQYGTRSLQKIIDTVDNEVQIDLIIKGFSQEFTSIEQVVT
                     LINDLNGNHVIQKCIFKFSPSKFGFIIDAIVEQNNIITISTHKHGCCVLQKLLSVCTL
                     QQIFKISVKIVQFLPGLINDQFGNYIIQFLLDIKELDFYLLAELFNRLSNELCQLSCL
                     KFSSNVVEKFIKKLFRIITGFIVNNNGGASQRTAVASDDVINASMNILLTTIDIFTVN
                     LNVLIRDNFGNYALQTLLDVKNYSPLLAYNKNSNAIGQNSSSTLNYGNFCNDFSLKIG
                     NLIVLTKELLPSIKTTSYAKKIKLKVKAYAEATGIPFTDISPQVTAMSHNNLQTINNE
                     NKNPHNKNSHNHNHNHNHNHAHNNNNNNNQKSHTRHFSLPANAYHRRSNSSVTNNFSN
                     QYAQDQKIHSPQQIMNFNQNAYPSMGAPSFNSQTNPPLVSHNSLQNFDNRQFANLMAH
                     PNSAAPIHSFSSSNITNVNPNVSRGFKQPGFMMNETDKINANHFSPYSNANSQNFNES
                     FVPQWNCTPRRIML"
```

```
     exon            4030. .4032
                     /gene="MPT5=HTR1"
                     /number=1
     intron          4033. .4672
                     /gene="MPT5=HTR1"
                     /number=1
     gene            4237. .4584
                     /gene="biB115"
     CDS             4237. .4584
                     /gene="biB115"
                     /note="putative protein"
                     /codon_start=1
                     /protein_id="CAA58662.1"
                     /db_xref="GI:1209470"
                     /db_xref="SGD:S000003145"
                     /db_xref="UniProt/Swiss-Prot:P53105"
                     /translation="MPIRIFSKEGLRERGASGKNEQKKKKKEKIRRSMKNLCRFVESY
                     KDFTLYGYRFHSKIKHKFCVLSSSFPVLFFCHPLSPLKSNFIDRFTLDSHASQATWTL
                     YMVVIVLFLSYRN"
     misc_feature    4643. .4650
                     /gene="MPT5=HTR1"
                     /note="splicing branchpoint"
     exon            4673. .7073
                     /gene="MPT5=HTR1"
                     /number=2
     gene            4745. .7249
                     /gene="bic834"
     CDS             4745. .7249
                     /gene="bic834"
                     /note="putative protein"
                     /codon_start=1
                     /protein_id="CAA58663.1"
                     /db_xref="GI:1209471"
                     /db_xref="GOA:P39016"
                     /db_xref="InterPro:IPR001313"
                     /db_xref="UniProt/Swiss-Prot:P39016"
                     /translation="MSYNHQPQLSINSVQSLLEPVTPPPLGQMNNKRNHQKAHSLDLS
                     GFNQFISSTQSPLALMNNTSTSNSANSFSPNPNAASNSTGLSASMANPPAILPLINEF
                     DLEMDGPRRKSSHDFTVVAPSNSGVNTSSLIMETPSSSVTPAASLRNFSNSNNAASKC
                     GVDNSSFGLSSSTSSSMVEISALPLRDLDYIKLATDQFGCRFLQKKLETPSESNMVRD
                     LMYEQIKPFFLDLILDPFGNYLVQKLCDYLTAEQKTLLIQTIYPNVFQISINQYGTRS
                     LQKIIDTVDNEVQIDLIIKGFSQEFTSIEQVVTLINDLNGNHVIQKCIFKFSPSKFGF
                     IIDAIVEQNNIITISTHKHGCCVLQKLLSVCTLQQIFKISVKIVQFLPGLINDQFGNY
                     IIQFLLDIKELDFYLLAELFNRLSNELCQLSCLKFSSNVVEKFIKKLFRIITGFIVNN
                     NGGASQRTAVASDDVINASMNILLTTIDIFTVNLNVLIRDNFGNYALQTLLDVKNYSP
                     LLAYNKNSNAIGQNSSSTLNYGNFCNDFSLKIGNLIVLTKELLPSIKTTSYAKKIKLK
                     VKAYAEATGIPFTDISPQVTAMSHNNLQTINNENKNPHNKNSHNHNHNHNHNHAHNNN
                     NNNNQKSHTRHFSLPANAYHRRSNSSVTNNFSNQYAQDQKIHSPQQIMNFNQNAYPSM
                     GAPSFNSQTNPPLVSHNSLQNFDNRQFANLMAHPNSAAPIHSFSSSNITNVNPNVSRG
                     FKQPGFMMNETDKINANHFSPYSNANSQNFNESFVPRMQYQTEGANWDSSLSMKSQHI
                     GQGPYNQVNMSRNASISNMPAMNTARTSDELQFTLP"
     intron          7074. .7911
                     /gene="MPT5=HTR1"
                     /number=2
     misc_feature    7835. .7842
                     /gene="MPT5=HTR1"
                     /note="splicing branchpoint"
     exon            7912. .7946
                     /gene="MPT5=HTR1"
                     /number=3
ORIGIN

Alignment Scores:
Pred. No.:              1.72e+03        Length:         7951
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x SCDNAGMH (1-7951)

Qy          1 HisAsnHisAsnHisAsn 6
                ||||||||||||||||||
Db       6569 CATAATCATAATCATAAT 6586

RESULT 360
AC015338
LOCUS       AC015338               7992 bp    DNA     linear   HTG 16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC015338
VERSION     AC015338.1  GI:6435997
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 7992)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10213177 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .7992
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
ORIGIN

Alignment Scores:
Pred. No.:              1.73e+03        Length:         7992
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AC015338 (1-7992)

Qy          1 HisAsnHisAsnHisAsn 6
                ||||||||||||||||||
Db       3702 CATAACCATAACCATAAC 3719

RESULT 361
AC013239/c
LOCUS       AC013239               8011 bp    DNA     linear   HTG 03-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC013239
VERSION     AC013239.1  GI:6223093
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 8011)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10214421 by the submitter.
            For further information on this sequence you may e-mail to
            fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .8011
```

```
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
ORIGIN

Alignment Scores:
Pred. No.:                1.74e+03        Length:          8011
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       14              Gaps:            0

US-09-858-332G-28 (1-6) x AC013239 (1-8011)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        184 CATAACCATAATCATAAC 167

RESULT 362
AF474341/c
LOCUS       AF474341               8167 bp    DNA     linear   INV 06-MAR-2002
DEFINITION  Dictyostelium discoideum ABC transporter ABCC.9 (abcC9) gene,
            complete cds.
ACCESSION   AF474341
VERSION     AF474341.1  GI:19172039
KEYWORDS    .
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 8167)
  AUTHORS   Anjard,C. and Loomis,W.F.
  TITLE     Evolution of the ABC transporters of Dictyostelium
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 8167)
  AUTHORS   Anjard,C. and Loomis,W.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JAN-2002) Biology Dept 0368, UCSD, 9500 Gilman drive,
            La Jolla, CA 92093, USA
FEATURES             Location/Qualifiers
     source          1. .8167
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="Ax4"
                     /db_xref="taxon:44689"
     gene            <320. .>4553
                     /gene="abcC9"
     mRNA            join(<320. .792,989. .>4553)
                     /gene="abcC9"
                     /product="ABC transporter ABCC.9"
     CDS             join(320. .792,989. .4553)
                     /gene="abcC9"
                     /note="protein family C"
                     /codon_start=1
                     /product="ABC transporter ABCC.9"
                     /protein_id="AAL85712.1"
                     /db_xref="GI:19172040"
                     /translation="MIRNKIKNKKSKSDYYSEISLEEEEDKEEEIKDKKVLSKEEIKE
                     NENKNEKIKQTQCPEDNASLFSKITFGWITKLIVKAYIKKSLDIKDIFDVPNYLKSNY
                     TSKFLTTEKQTFKRNTKYSLILNIYINFILLKNKKLILIQFLRVLFTLLSPLILKLFI
                     EFTQRIDSEKSIIEGIIYCGLLFFSSFCSSLVDEYLVWFGMITSSQVKSCLTCLIFEK
                     SLKLSNSAKVKYNAAKITNLISIDVDNFSNFFWSNSVEIFFQPFQLIFLLILLIPEVG
                     WSGFVGTAVILISFPINSYFGKKTSDYYEKLLKYTDKRVSTTSEFINGIRFIKMYAWE
                     SLFLKKIEKQRKQELKILLERGLFWLGQIIIINTNSTFVFVATMVTYSLSGNKMKLET
                     AFTAMNILDSIRILLIVMPYCYYSIMELIPSNKRIEKFLSTLEIQQNLQTTKNSNTIS
                     INNGTFKWKDENFEETDDDDDHDDDDDEGEKEIEGKEEKEEINNFIFENINFKAPIGK
                     LTMICSPVGSGKTSLINALIGEIEKVNGEINGVPENISFTSQQPFLLSTSLRENILFG
                     KELDIEYYKQVLDACCLVSDLTQLSALDLTEIGEGGINLSGGQKQRVSLARALYSNSD
                     FILMDEPLSAVDPEVANHLFEKCIQGMMKNKTRILVTHQIQFIPYADHIVIIKDGKII
                     QQGTYKELKDENGIDFESIIKTKNSNLNLNKTIKDEEEKEDKEGEEKEEDKKEMILKLI
                     KINYSDELKEKAKLLVEEDRNEGEVSFKTYKEYFYYGSSNLFLFITLLVFLIGLIVNR
                     LSDFWLTIWTEQSFKEKSQTFYIICFVSIFIVSLILIFIRYSLVAKIGFKSSKKLHDL
                     LINSISKSPIQFFDSNPSGRILNRFSNDISDIDTSMIDMFSDTLEYIFAIVVGIFSII
                     YINPMTIVPFLILSMFYYQIFNIYRVSVRELNRCKSISQSPIISFLSECCNGLSTIRA
                     FKQQSRFIGIMNDNIDKNLKCEFASFAVEMWISIRLELLSSIIVFIVSLFSLFNGYSN
                     SALSILSVSTARSITSYLKSACRQMVDLERKMNSFERVNNYIKLPSEGSGSGVKSMEF
                     ISEKDLLNWPEKGNIQFNNVQVNYNSNSVPSLKDITFSVESNEKIGIVGRTGAGKSTI
                     ANCLFRLVECSKGSILIDGIDIKTIDLNKLRGSLGIVPQEPWLFSGTIRSNIDPLNQY
                     DDEMIWNYLEMVKLKKLIIEMPLKLNSKIHENGNTTLSYGQKQLLCLCRCLIKNPKLI
                     IMDEATSSVDFQTAETIKSVINENLVNNTILTIAHRLDIIIDSDKIAVIDNSKLIEFE
                     NPKNLINSNSKFRKIVNFQTKIQTN"
ORIGIN

Alignment Scores:
Pred. No.:                1.77e+03        Length:          8167
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       2               Gaps:            0

US-09-858-332G-28 (1-6) x AF474341 (1-8167)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       7407 CATAACCATAATCATAAT 7390

RESULT 363
CDR131945
LOCUS       CDR131945              8724 bp    DNA     linear   MAM 07-JAN-2000
DEFINITION  Camelus dromedarius IGG2a gene, alternative transcripts.
ACCESSION   AJ131945
VERSION     AJ131945.1  GI:6687180
KEYWORDS    alternative splicing; constant region; IGG2a gene; immunoglobulin
            heavy chain; membrane-bound form; secreted form.
SOURCE      Camelus dromedarius (Arabian camel)
  ORGANISM  Camelus dromedarius
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Tylopoda;
            Camelidae; Camelus.
REFERENCE   1
  AUTHORS   Nguyen,V.K., Hamers,R., Wyns,L. and Muyldermans,S.
  TITLE     Loss of splice consensus signal is responsible for the removal of
            the entire C(H)1 domain of the functional camel IGG2A heavy-chain
            antibodies
  JOURNAL   Mol. Immunol. 36 (8), 515-524 (1999)
   PUBMED   10475606
REFERENCE   2  (bases 1 to 8724)
  AUTHORS   Nguyen,K.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-1999) Nguyen K.V., ULTR, Insitute of
            biotechnology, Vrije Universiteit Brussel, Paardenstraat 65, Sint
            Genesius Rode, B-1640, BELGIUM
FEATURES             Location/Qualifiers
     source          1. .8724
                     /organism="Camelus dromedarius"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9838"
                     /clone="ch51666E2"
                     /tissue_type="liver"
                     /germline
     gene            1. .3863
                     /gene="IGG2a"
     5'UTR           1. .319
                     /gene="IGG2a"
     exon            320. .613
                     /gene="IGG2a"
                     /number=1
     intron          614. .929
                     /gene="IGG2a"
                     /number=1
     CDS             join(<930. .1034,1158. .1487,1571. .1891,3265. .3398,
                     3780. .3863)
                     /gene="IGG2a"
                     /note="alternative splice variant
                     membrane-bound form"
```

```
                     /codon_start=3
                     /product="immunoglobulin heavy chain constant region"
                     /protein_id="CAB64865.1"
                     /db_xref="GI:6687182"
                     /translation="PKIPQPQPKPQPQPQPQPKPQPKPEPECTCPKCPAPELLGGPSV
                     FIFPPKPKDVLSISGRPEVTCVVVDVGQEDPEVSFNWYIDGVEVRTANTKPKEEQFNS
                     TYRVVSVLTIQHQDWLTGKELKCKVNNKALPAPIERTISKAKGQTREPQVYTLAPHRE
                     ELAKDTVSITCLVIGFYPADINVEWQRNGRPESEGAYATTLPQLDNDGTYFLYSKLSV
                     GKNTWQQGETFTCVVMHEALHNHSTQKSITQSSEPLLEEESCAEAQSGELDGLWTTIS
                     IFITLFLLSVCYSATVTLFKVKWIFSSVVELKRTIVPDYRNMIGQGS"
C_region             join(<930. .1034,1158. .1487,1571. .1891,3265. .3398,
                     3780. .3863)
                     /gene="IGG2a"
                     /product="immunoglobulin heavy chain constant region"
                     /note="alternative splice variant
                     membrane-bound form"
CDS                  join(<930. .1034,1158. .1487,1571. .1899)
                     /gene="IGG2a"
                     /note="alternative splice transcript
                     secreted form"
                     /codon_start=3
                     /product="immunoglobulin heavy chain constant region"
                     /protein_id="CAB64864.1"
                     /db_xref="GI:6687181"
                     /translation="PKIPQPQPKPQPQPQPQPKPQPKPEPECTCPKCPAPELLGGPSV
                     FIFPPKPKDVLSISGRPEVTCVVVDVGQEDPEVSFNWYIDGVEVRTANTKPKEEQFNS
                     TYRVVSVLTIQHQDWLTGKELKCKVNNKALPAPIERTISKAKGQTREPQVYTLAPHRE
                     ELAKDTVSITCLVIGFYPADINVEWQRNGRPESEGAYATTLPQLDNDGTYFLYSKLSV
                     GKNTWQQGETFTCVVMHEALHNHSTQKSITQSSGK"
C_region             join(<930. .1034,1158. .1487,1571. .1899)
                     /gene="IGG2a"
                     /product="immunoglobulin heavy chain constant region"
                     /note="alternative splice transcript
                     secreted form"
exon                 930. .1034
                     /gene="IGG2a"
                     /number=2
intron               1035. .1157
                     /gene="IGG2a"
                     /number=2
exon                 1158. .1487
                     /gene="IGG2a"
                     /number=3
intron               1488. .1570
                     /gene="IGG2a"
                     /number=3
exon                 1571. .1899
                     /gene="IGG2a"
                     /note="number 4a"
exon                 1571. .1891
                     /gene="IGG2a"
                     /note="number 4b"
intron               1892. .3264
                     /gene="IGG2a"
                     /note="number 4b"
intron               1897. .3264
                     /gene="IGG2a"
                     /note="number 4a"
polyA_signal         2000. .2005
                     /gene="IGG2a"
polyA_site           2033
                     /gene="IGG2a"
exon                 3265. .3398
                     /gene="IGG2a"
                     /note="M1 membrane exon
                     putative"
                     /number=5
intron               3399. .3779
                     /gene="IGG2a"
                     /number=5
exon                 3780. .3863
                     /gene="IGG2a"
                     /note="M2 membrane exon"
                     /number=6
     polyA_signal    4879. .4884
     polyA_signal    6640. .6645
ORIGIN

Alignment Scores:
Pred. No.:                1.88e+03        Length:          8724
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       4               Gaps:            0

US-09-858-332G-28 (1-6) x CDR131945 (1-8724)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        960 CACAACCACAACCACAAC 977

RESULT 364
AY191010
LOCUS       AY191010                8867 bp    DNA     linear   PLN 26-FEB-2003
DEFINITION  Zea mays Bz(Dp26)-m1 bz locus tandem duplication sequence.
ACCESSION   AY191010
VERSION     AY191010.1  GI:28569515
KEYWORDS    .
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 8867)
  AUTHORS   Dooner,H.K. and Martinez-Ferez,I.M.
  TITLE     Germinal excisions of the maize transposon activator do not
            stimulate meiotic recombination or homology-dependent repair at the
            bz locus
  JOURNAL   Genetics 147 (4), 1923-1932 (1997)
   PUBMED   9409847
REFERENCE   2  (bases 1 to 8867)
  AUTHORS   Dooner,H.K. and Zheng,Z.
  TITLE     De novo origin of a tandem duplication in maize by improper
            resolution of a recombination intermediate
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 8867)
  AUTHORS   Zheng,Z., Martinez-Ferez,I.M. and Dooner,H.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2002) Waksman Institute, Rutgers University, Hoes
            Lane, Piscataway, NJ 08855, USA
FEATURES             Location/Qualifiers
     source          1. .8867
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="W22"
                     /db_xref="taxon:4577"
                     /chromosome="9"
                     /map="9S; between stk1 and stc1"
                     /note="tandem duplication of bz locus
                     genotype: Bz(Dp26)-m1; Ac-mutable allele"
     misc_feature    1. .6905
                     /note="proximal member of Dp26 duplication; carries an
                     Ac-mutable hybrid bz allele produced by intragenic
                     recombination between the bz-m1 mutant from Bz-McC found
                     in GenBank Accession Numbers AF037598 and X07940 and the
                     bz-E4 mutant from Bz-W22 found in GenBank Accession
                     Numbers AY167674 and X07937"
     repeat_region   1308. .1315
                     /note="target site direct repeat"
     repeat_region   1316. .5880
                     /transposon="Ac"
     repeat_region   5881. .5888
                     /note="target site direct repeat"
     misc_recomb     6906. .6911
                     /note="duplication junction"
```

```
     misc_feature    6912. .8867
                     /note="distal member of Dp26 duplication; carries a mutant
                     bz-E4 allele"
     variation       8182
                     /note="E4 mutant site"
ORIGIN

Alignment Scores:
Pred. No.:                1.91e+03        Length:         8867
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       15              Gaps:           0

US-09-858-332G-28 (1-6) x AY191010 (1-8867)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       2735 CACAACCACAACCACAAC 2752

RESULT 365
AF411931
LOCUS       AF411931                8979 bp    DNA     linear   INV 04-AUG-2003
DEFINITION  Plasmodium falciparum strain Dd2 normocyte-binding protein 1 (NBP1)
            gene, complete cds.
ACCESSION   AF411931
VERSION     AF411931.2  GI:33414604
KEYWORDS    .
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 8979)
  AUTHORS   Rayner,J.C., Vargas-Serrato,E., Huber,C.S., Galinski,M.R. and
            Barnwell,J.W.
  TITLE     A Plasmodium falciparum Homologue of Plasmodium vivax Reticulocyte
            Binding Protein (PvRBP1) Defines a Trypsin-resistant Erythrocyte
            Invasion Pathway
  JOURNAL   J. Exp. Med. 194 (11), 1571-1582 (2001)
   PUBMED   11733572
REFERENCE   2  (bases 1 to 8979)
  AUTHORS   Rayner,J.C., Barnwell,J.W., Galinski,M.R., Tran,T., Huber,C.S. and
            Corredor,V.
  TITLE     Differing geographic diversity among members of the Reticulocyte
            Binding Like (RBL) gene family of the two major human malaria
            parasites, Plasmodium falciparum and Plasmodium vivax
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 8979)
  AUTHORS   Rayner,J.C., Galinski,M.R. and Barnwell,J.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-AUG-2001) Division of Parasitic Diseases, Centers for
            Disease Control and Prevention, 4770 Buford Highway, MS F-13,
            Chamblee, GA 30341, USA
REFERENCE   4  (bases 1 to 8979)
  AUTHORS   Rayner,J.C., Barnwell,J.W. and Galinski,M.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-2002) Division of Parasitic Diseases, Centers for
            Disease Control and Prevention, 4770 Buford Highway, MS F-13,
            Chamblee, GA 30341, USA
  REMARK    Sequence update by submitter
COMMENT     On Aug 4, 2003 this sequence version replaced gi:17352159.
FEATURES             Location/Qualifiers
     source          1. .8979
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /strain="Dd2"
                     /db_xref="taxon:5833"
     gene            <8. .>8979
                     /gene="NBP1"
     mRNA            join(<8. .62,161. .>8979)
                     /gene="NBP1"
                     /product="normocyte-binding protein 1"
     CDS             join(8. .62,161. .8979)
                     /gene="NBP1"
                     /note="similar to Plasmodium vivax reticulocyte binding
                     protein 1"
                     /codon_start=1
                     /product="normocyte-binding protein 1"
                     /protein_id="AAL38220.2"
                     /db_xref="GI:33414605"
                     /translation="MQRWIFCNIVLHILIYLAEFSHEQESYSSNEKIRKDYSDDNNYE
                     PTPSYEKRKKEYGKDESYIKNYRGNNFSYDLSKNSSIFLHMGNGSNSKTLKRCNKKKN
                     IKTNFLRPIEEEKTVLNNYVYKGVNFLDTIKRNDSSYKFDVYKDTSFLKNREYKELIT
                     MQYDYAYLEATKEVLYLIPKDKDYHKFYKNELEKILFNLKDSLKLLREGYIQSKLEMI
                     RIHSDIDILNEFHQGNIINDNYFNNEIKKKKEDVEKYIREYNLYIYKYENQLKIEIQK
                     LTNEVSINLNKSTCEKNCYYYILKLEKYKNIIKDKINKWTDLPEIYIDDKSFSYTFLK
                     DVINNKIDIYKTISSFISTQKQLYYFEYIYIMNKNTLNLLSYNIQKTDINSSSKYTYT
                     KSHFLKDNHILLSKYYTAKFIDILNKTYYYNLYKNKILLFNKYIIKLRNDLKEYAFKS
                     IQFIQDKIKKHKDELSIENILQEVNNIYIKYDTSINEISKYNNLIINTDLQIVQQKLL
                     EIKQKKNDITHKVQLINHIYKNIHDEILNKKNNEITKIIINNIKDHKKDLQDLLLFIQ
                     QIKQYNILTDHNITQCNNYYKEIIKMKEDINHIHIYIQPILNNLHTLKQVQNNKIKYE
                     EHIKQILQKIYDKKESLKKIILLKDEAQLDITLLDDLIQKQTQTQKQTLIQNNETIQL
                     ISGQEDKHESNPFNHIQTYIQQKDTQNKNIQNLLKSLYNGNINTFIDTISKYILKQKD
                     IELTQHVYTDEKINDYLEEIKNEQNKIDKTIDDIKIQETLKQITHIVNNIKTIKKDLL
                     KEFIQHLIKYMNERYQNMQQGYNNLTNYINQYEEENNNMKQYITTIRNIQKIYYDNIY
                     AKEKEIRSGQYYKDFITSRKNIYNIRENISKNVDMIKNEEKKKIQNCVDKYNSIKQYV
                     KMFKNGDTQDENNNNNNDIYDKLIVPLDSIKQNIDKYNTEHNFITFTNKINTHNKKNQ
                     EMMEEFIYVYKRLKILKILNISLKACEKNNKSINTLNDKTQELKKIVTHEIDLLQKDI
                     LTSQISNKNVLLLNDLLKEIEQYIIDIHKLKKKSNDLFTYYEQSKNYFYFKNKKDNFD
                     IQKTINKMNEWLAIKNYINEINKNYQTLYEKKINVLLHNSKSYVQYFYDHIINLILQK
                     KNYLENTLKTKIQDNEHSLYALQQNEEYQKVKNEKDQNEIKKIKQLIEKNKNDILTYE
                     NNIEQIEQKNIELKTNAQNKDDQIVNTLNEVKKKIIYTYEKVDNQISNVLKNYEEGKV
                     EYDKNVVQNVNDADDTNDIDEINDIDEINDIDEINDIDEIKDIDHIKHFDDTKHFDDI
                     YHADDTRDEYHIALSNYIKTELRNINLQEIKNNIIKIFKEFESAHKEIKKESKQINKE
                     FTKMDVVINQLRDIDRQMLDLYKELDEKYSEFNKTKIEEINNIRENINNVEIWYEKNI
                     IEYFLRYMNDQKDKAAKYMENIDTYKNNIEIISKQINPENYVETLKKTNMYSYVEKAN
                     DLFYKQINNIIINSNQLKNEAFTIDELQNIQKNRKNLLTKKQQIIQYTNEIENIFNEI
                     KNINNILVLTNYKSILQDISQNINHVSIYTEQLHNLYIKLEEEKEQMKTLYHKSNVLH
                     NQINFNEDAFINNLLINIEKIKNDITHIKEKTNIYMIDVNKSKNNAQLYFHNTLRGNE
                     KIEYLKNLKNSTNQQITLQELKQVQENVEKVKDIYNQTIKYEEKIKKNYHIITDYENK
                     INDILHNSFIKQINMESSNNKKQTKQIIDIINDKTFEEHIKTSKTKINMLKEQSQMKH
                     IDKTLLNEQALKLFVDINSTNNNLDNMLSEINSIQNNIHTYIQEANKSFDKFKIICDQ
                     NVNDLLNKLSLGDLNYMNHLKNLQNEIRNMNLEKNFMLDKSKKIDEEEKKLDILKVNI
                     SNINNSLDKLKKYYEEALFQKVKEKAEIQKENIEKIKQEINTLSDVFKKPFFFIQLNT
                     DSSQHEKDINNNVETYKNNIDEIYNVFIQSYNLIQKYSSEIFSSTLNYIQTKEIKEKS
                     IKEQNQLNQNEKEASVLLKNIKINETIKLFKQIKNERQNDVHNIKEDYNLLQQYLNYM
                     KNEMEQLKKYKNDVHMDKNYVENNNGEKEKLLKETISSYYDKINNINNKLYIYKNKED
                     TYFNNMIKVSEILNIIIKKKQQNEQRIVINAEYDSSLINKDEEIKKEINNQIIELNKH
                     NENISIIFKDIQNIKKQSQDIITNMNDMHKSTILLVDIIQKKEEALNKQKNILRNIDN
                     ILNKKENIIDKVIKCNCDDYKDILIQNETEYQKLQNINHTYEEKKKSIDILKIKNIKQ
                     KNIQEYKNKLEQMNTIINQSIEQHVFINADILQNEKIKLEEIIKNLDILDEQIMTYHN
                     SIDELYKLGIQCDNHLITTISVVVNKNTTKIMIHIKKQKEDIQKINNYIQTNYNIINE
                     EALQFHRLYGHNLISEDDKNNLVHIIKEQKNIYTQKEIDISKIIKHVKKGLYSLNEHD
                     MNHDTHMNIINEHINNNILQPYTQLINMIKDIDNVFIKIQNNKFEQIQKYIEIIKSLE
                     QLNKNINTDNLNKLKDTQNKLINIETEMKHKQKQLINKMNDIEKDNITDQYMHDVQQN
                     IFEPITLKMNEYNTLLNDNHNNNINNNEHQFNHLNSLHTKIFSHNYNKEQQQEYITNIM
                     QRIDVFINDLDTYQYEYYFYEWNQEYKQIDKNKINQHINNIKNNLIHVKKQFEHTLEN
                     IKNNENIFDNIQLKKKDIDDIIININNTKETYLKELNKKKMLQNKKKVDEKSEINNHH
                     TLQHDNQNVEQKNKIKDHNLITKPNNNSSEESHQNEQMKEQNKNILEKQTRNIKPHHV
                     HNHNHNHNQNQKDSTKLQEQDISTHKLHNTIHEQQSKDNHQGNREKKQKNGNHERMYF
                     ASGIVVSILFLSSLGFVINSKNNKQEYDKEQEKQQQNDFVCDNNKMDDKSTQKYGRNQ
                     EEVMEISFDNDYI"
ORIGIN

Alignment Scores:
Pred. No.:                1.94e+03        Length:         8979
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       2               Gaps:           0

US-09-858-332G-28 (1-6) x AF411931 (1-8979)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
```

```
Db       8590 CATAATCATAATCATAAT 8607

RESULT 366
AF411932
LOCUS       AF411932               9003 bp    DNA     linear   INV 04-AUG-2003
DEFINITION  Plasmodium falciparum strain Malayan Camp K- normocyte-binding
            protein 1 (NBP1) gene, complete cds.
ACCESSION   AF411932
VERSION     AF411932.2  GI:33414606
KEYWORDS    .
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 9003)
  AUTHORS   Rayner,J.C., Vargas-Serrato,E., Huber,C.S., Galinski,M.R. and
            Barnwell,J.W.
  TITLE     A Plasmodium falciparum Homologue of Plasmodium vivax Reticulocyte
            Binding Protein (PvRBP1) Defines a Trypsin-resistant Erythrocyte
            Invasion Pathway
  JOURNAL   J. Exp. Med. 194 (11), 1571-1582 (2001)
   PUBMED   11733572
REFERENCE   2  (bases 1 to 9003)
  AUTHORS   Rayner,J.C., Barnwell,J.W., Galinski,M.R., Tran,T., Huber,C.S. and
            Corredor,V.
  TITLE     Differing geographic diversity among members of the Reticulocyte
            Binding Like (RBL) gene family of the two major human malaria
            parasites, Plasmodium falciparum and Plasmodium vivax
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 9003)
  AUTHORS   Rayner,J.C., Galinski,M.R. and Barnwell,J.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-AUG-2001) Division of Parasitic Diseases, Centers for
            Disease Control and Prevention, 4770 Buford Highway, MS F-13,
            Chamblee, GA 30341, USA
REFERENCE   4  (bases 1 to 9003)
  AUTHORS   Rayner,J.C., Barnwell,J.W. and Galinski,M.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-2002) Division of Parasitic Diseases, Centers for
            Disease Control and Prevention, 4770 Buford Highway, MS F-13,
            Chamblee, GA 30341, USA
  REMARK    Sequence update by submitter
COMMENT     On Aug 4, 2003 this sequence version replaced gi:17352161.
FEATURES             Location/Qualifiers
     source          1. .9003
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /strain="Malayan Camp K-"
                     /db_xref="taxon:5833"
     gene            <8. .>9003
                     /gene="NBP1"
     mRNA            join(<8. .62,161. .>9003)
                     /gene="NBP1"
                     /product="normocyte-binding protein 1"
     CDS             join(8. .62,161. .9003)
                     /gene="NBP1"
                     /note="similar to Plasmodium vivax reticulocyte binding
                     protein 1"
                     /codon_start=1
                     /product="normocyte-binding protein 1"
                     /protein_id="AAL38221.2"
                     /db_xref="GI:33414607"
                     /translation="MQRWIFCNIVLHILIYLAEFSHEQESYSSNEKIRKDYSDDNNYE
                     PTPSYEKRKKEYGKDESYIKNYRGNNFSYDLSKNSSIFLHMGNGSNSKTLKRCNKKKN
                     IKTNFLRPIEEEKTVLNNYVYKGVNFLDTIKRNDSSYKFDVYKDTSFLKNREYKELIT
                     MQYDYAYLEATKEVLYLIPKDKDYHKFYKNKLEKILFNLKDSLKLLREGYIQSKLEMI
                     RIDSDIDILNEFHQGNIINDNYFNNEIKKKKEDVEKYIREYNLYIYKYENQLKIEIQK
                     LTNEVSINLNKSTCEKNCYYYILKLEKYKNIIKDKINKWKDLPEIYIDDKSFSYTFLK
                     DVINNKIDIYKTISSFISTQKQLYYLEYIYIMNKNTLNLLSYNIQKTDINSSSKYTYT
                     KSHFLKDNHILLSKYYTAKFIDILNKTYYYNLYKNKILLFNKYIIKLRNDLKEYAFKS
                     IQFIQDKIKKHKDELSIENILQEVNNIYIKYDTSINEISKYNNLIINTDLQIVQQKLL
                     EIKQKKNDITHKVQLINHIYKNIHDEILNKKNNEITKIIINNIKDHKKDLQDLLLFIQ
                     QIKQYNILTDHNITQCNNYYKEIIKMKEDINHIHIYIQPILNNLHTLKQVQNNKIKYE
                     EHIKQILQKIYDKKESLKKIILLKDEAQLDITLLDDLIQKQTQTQTQTQTQKQTLIQN
                     NETIQLISGQEDKHESNPFNHIQTYIQQKDTQNKNIQNLLKSLYNGNINTFIDTISKY
                     ILKQKDIELTQHVYTDEKINDYLEEIKNEQNKIDKTIDDIKIQETLKQITHIVNNIKT
                     IKKDLLKEFIQHLIKYMNERYQNMQQGYNNLTNYINQYEEENNNMKQYITTIRNIQKI
                     YYDNIYAKEKEIRSGQYYKDFITSRKNIYNIRENISKNVDMIKNEEKKKIQNCVDKYN
                     SIKQYVKMFKNGDTQDENNNNNNDIYDKLIVPLDSIKQNIDKYNTEHNFITFTNKINT
                     HNKKNQEMMEEFIYVYKRLKIFKILNISLKACEKNNKSINTLNDKTQELKKIVTHEID
                     LLQKDILTSQISNKNVLLLNDLLKEIEQYIIDVHKLKKKSNDLFTYYEQSKNYFYFKN
                     KKDNFDIQKTINKMNEWLAIKNYINEINKNYQTLYEKKINVLLHNSKSYVQYFYDHII
                     NLILQKKNYLENTLKTKIQDNEHSLYALQQNEEYQKVKNEKDQNEIKKIKQLIEKNKN
                     DILTYENNIEQIEQKNIELKTNAQNKDDQIVNTLNEVKKKIIYTYEKVDNQISNVLKN
                     YEEGKVEYDKNVVQNVNDADDTNDIDEINDIDEINDIDEINDIDEIKDIDHIKHFDDT
                     KHFDDIYHADDTRDEYHIALSNYIKTELRNINLQEIKNNIIKIFKEFESAHKEIKKES
                     KQINKEFTKMDVVINQLRDIDRQMLDLYKELDEKYSEFNKTKIEEINNIRENINNVEI
                     WYEKNIIEYFLRHMNDQKDKAAKYMENIDTYKNNIEIISKQINPENYVETLNKSNMYS
                     YVEKANDLFYKQINNIIINSNQLKNEAFTIDELQNIQKNRKNLLTKKQQIIQYTNEIE
                     NIFNEIKNINNILVLTNYKSILQDISQNINHVSIYTEQLHNLYIKLEEEKEQMKTLYH
                     KSNVLHNQINFNEDAFINNLLINIEKIKNDITHIKEKTNIYMIDVNKSKNNAQLYFHN
                     TLRGNEKIEYLKNLKNSTNQQITLQELKQVQENVEKVKDIYNQTIKYEEKIKTNYHII
                     TDYENKINDILHNSFIKQINMESSNNKKQTKQIIDIINDKTFEEHIKTSKTKINMLKE
                     QSQMKHIDKTLLNEQALKLFVDINSTNNNLDNMLSEINSIQNNIHTYIQEANKSFDKF
                     KIICDQNVNDLLNKLSLGDLNYMNHLKNLQNEIRNMNLEKNFMLDKSKKIDEEEKKLD
                     ILKVNISNINNSLDKLKKYYEEALFQKVKENAEIQKENIEKIKQEINTLSDVFKKPFF
                     FIQLNTDSSQHEKDINNNVETYKNNIDEIYNVFIQSYNLIQKYSSEIFSSTLNYIQTK
                     EIKEKSIKEQNQLNQNEKEASVLLKTIKINETIKLFKQIKNERQNDVHNIKEDYNLLQ
                     QYLNYMKNEMEQLKKYKIDVHMDKNYVENNNGEKEKLLKETISSYYDKINNINNKLYI
                     YKNKEDTYFNNMIKVSEILNIIIKKKQQNEQRIVINAEYDSSLINKDEEIKKEINNQI
                     IELNKHNENISNIFKDIQNIKKQSQDIITNMNDMHKSTILLVDIIQKKEEALNKQKNI
                     LRNIDNILNKKENIIDKVIKCNCDDYKDILIQNETEYQKLQNINHTYEEKKKSIDILK
                     IKNIKQKNIQEYKNKLEQMNTIINQSIEQHVFINADILQNEKIKLEEIIKNLDILDEQ
                     IMTYHNSIDELYKLGIQCDNHLITTISVVVNKNTTKIMIHIKKQKEDIQKINNYIQTN
                     YNIINEEALQFHRLYGHNLISEDDKNNLVHIIKEQKNIYTQKEIDISKIIKHVKKGLY
                     SLNEHDMNHDTHMNIINEHINNNILQPYTQLINMIKDIDNVFIKIQNNKFEQIQKYIE
                     IIKSLEQLNKNINTDNLNKLKDTQNKLINIETEMKHKQKQLINKMNDIEKDNITDQYM
                     HDVQQNIFEPITLKMNEYNTLLNDNHNNNINNEHQFNHLNSLHTKIFSHNYNKEQQQE
                     YITNIMQRIDVFINDLDTYQYEYYFYEWNQEYKQIDKNKINQHINNIKNNLIHVKKQF
                     EHTLENIKNNENIFDNIQLKKKDIDDIIININNTKETYLKELNKKKMLQNKKKVDEKS
                     EINNHHTLQHDNQNVEQKNKIKDHNLITKPNNNSSEESHQNEQMKEQNKNILEKQTRN
                     IKPHHVHNHNHNHNQNQNQKDSTKLQEQDISTHKLHNTIHEQQSKDNHQGNREKKQKN
                     GNHERMYFASGIVVSILFLSSLGFVINSKNNKQEYDKEQEKQQQNDFVCDNNKMDDKS
                     TQKYGRNQEEVMEISFDNDYI"
ORIGIN

Alignment Scores:
Pred. No.:                1.94e+03        Length:          9003
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       2               Gaps:            0

US-09-858-332G-28 (1-6) x AF411932 (1-9003)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       8608 CATAATCATAATCATAAT 8625

RESULT 367
AF533700
LOCUS       AF533700               9022 bp    DNA     linear   INV 08-AUG-2003
DEFINITION  Plasmodium falciparum 3D7 normocyte-binding protein 1 (NBP1) gene,
            complete cds.
ACCESSION   AF533700
VERSION     AF533700.1  GI:33329883
KEYWORDS    .
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 9022)
  AUTHORS   Rayner,J.C., Barnwell,J.W., Galinski,M.R., Tran,T., Huber,C.S. and
            Corredor,V.
  TITLE     Differing geographic diversity among members of the Reticulocyte
            Binding Like (RBL) gene family of the two major human malaria
            parasites, Plasmodium falciparum and Plasmodium vivax
```

```
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 9022)
  AUTHORS   Rayner,J.C., Barnwell,J.W., Galinski,M.R., Tran,T., Huber,C.S. and
            Corredor,V.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JUL-2002) Biology and Diagnostics Branch, Division of
            Parasitic Diseases, National Center for Infectious Diseases, Center
            for Disease Control and Prevention, 4770 Buford Highway, MS F-13,
            Atlanta, GA 30341, USA
FEATURES             Location/Qualifiers
     source          1. .9022
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /isolate="3D7"
                     /db_xref="taxon:5833"
     gene            <8. .>8418
                     /gene="NBP1"
                     /note="PfNBP1"
     mRNA            join(<8. .62,161. .>8418)
                     /gene="NBP1"
                     /product="normocyte-binding protein 1"
     CDS             join(8. .62,161. .8418)
                     /gene="NBP1"
                     /codon_start=1
                     /product="normocyte-binding protein 1"
                     /protein_id="AAQ10315.1"
                     /db_xref="GI:33329884"
                     /translation="MQRWIFCNIVLHILIYLAEFSHEQESYSSNEKIRKDYSDDNNYE
                     PTPSYEKRKKEYGKDESYIKNYRGNNFSYDLSKNSSIFLHMGNGSNSKTLKRCNKKKN
                     IKTNFLRPIEEEKTVLNNYVYKGVNFLDTIKRNDSSYKFDVYKDTSFLKNREYKELIT
                     MQYDYAYLEATKEVLYLIPKDKDYHKFYKNELEKILFNLKDSLKLLREGYIQSKLEMI
                     RIHSDIDILNEFHQGNIINDNYFNNEIKKKKEDMEKYIREYNLYIYKYENQLKIKIQK
                     LTNEVSINLNKSTCEKNCYNYILKLEKYKNIIKDKINKWKDLPEIYIDDKSFSYTFLK
                     DVINNKIDIYKTISSFISTQKQLYYFEYIYIMNKNTLNLLSYNIQKTDINSSSKYTYT
                     KSHFLKDNHILLSKYYTAKFIDILNKTYYYNLYKNKILLFNKYIIKLRNDLKEYAFKS
                     IQFIQDKIKKHKDELSIENILQEVNNIYIKYDTSINEISKYNNLIINTDLQIVQQKLL
                     EIKQKKNDITHKVQLINHIYKNIHDEILNKKNNEITKIIINNIKDHKKDLQDLLLFIQ
                     QIKQYNILTDHKITQCNNYYKEIIKMKEDINHIHIYIQPILNNLHTLKQVQNNKIKYE
                     EHIKQILQKIYDKKESLKKIILLKDEAQLDITLLDDLIQKQTKKQTQTQTQTQKQTLI
                     QNNETIQLISGQEDKHESNPFNHIQTYIQQKDTQNKNIQNLLKSLYNGNINTFIDTIS
                     KYILKQKDIELTQHVYTDEKINDYLEEIKNEQNKIDKTIDDIKIQETLKQITHIVNNI
                     KTIKKDLLKEFIQHLIKYMNERYQNMQQGYNNLTNYINQYEEENNNMKQYITTIRNIQ
                     KIYYDNIYAKEKEIRSGQYYKDFITSRKNIYNIRENISKNVDMIKNEEKKKIQNCVDK
                     YNSIKQYVKMLKNGDTQDENNNNNNDIYDKLIVPLDSIKQNIDKYNTEHNFITFTNKI
                     NTHNKKNQEMMEEFIYAYKRLKILKILNISLKACEKNNKSINTLNDKTQELKKIVTHE
                     IDLLQKDILTSQISNKNVLLLNDLLKEIEQYIIDVHKLKKKSNDLFTYYEQSKNYFYF
                     KNKKDNFDIQKTINKMNEWLAIKNYINEINKNYQTLYEKKINVLLHNSKSYVQYFYDH
                     IINLILQKKNYLENTLKTKIQDNEHSLYALQQNEEYQKVKNEKDQNEIKKIKQLIEKN
                     KNDILTYENNIEQIEQKNIELKTNAQNKDDQIVNTLNEVKKKIIYTYEKVDNQISNVL
                     KNYEEGKVEYDKNVVQNVNDADDTNDIDEINDIDEINDIDEINDIDEINDIDEIKDID
                     HIKHFDDTKHFDDIYHADDTRDEYHIALSNYIKTELRNINLQEIKNNIIKIFKEFKSA
                     HKEIKKESEQINKEFTKMDVVINQLRDIDRQMLDLYKELDEKYSEFNKTKIEEINNIR
                     ENINNVEIWYEKNIIEYFLRHMNDQKDKAAKYMENIDTYKNNIEIISKQINPENYVET
                     LNKSNMYSYVEKANDLFYKQINNIIINSNQLKNEAFTIDELQNIQKNRKNLLTKKQQI
                     IQYTNEIENIFNEIKNINNILVLTNYKSILQDISQNINHVSIYTEQLHNLYIKLEEEK
                     EQMKTLYHKSNVLHNQINFNEDAFINNLLINIEKIKNDITHIKEKTNIYMIDVNKSKN
                     NAQLYFHNTLRGNEKIEYLKNLKNSTNQQITLQELKQVQENVEKVKDIYNQTIKYEEK
                     IKKNYHIITDYENKINDILHNSFIKQINMESSNNKKQTKQIIDIINDKTFEEHIKTSK
                     TKINMLKEQSQMKHIDKTLLNEQALKLFVDINSTNNNLDNMLSEINSIQNNIHTYIQE
                     ANKSFDKFKIICDQNVNDLLNKLSLGDLNYMNHLKNLQNEIRNMNLEKNFMLDKSKKI
                     DEEEKKLDILKVNISNINNSLDKLKKYYEEALFQKVKEKAEIQKENIEKIKQEINTLS
                     DVFKKPFFFIQLNTDSSQHEKDINNNVETYKNNIDEIYNVFIQSYNLIQKYSSEIFSS
                     TLNYIQTKEIKEKSIKEQNQLNQNEKEASVLLKNIKINETIKLFKQIKNERQNDVHNI
                     KEDYNLLQQYLNYMKNEMEQLKKYKNDVHMDKNYVENNNGEKEKLLKETISSYYDKIN
                     NINNKLYIYKNKEDTYFNNMIKVSEILNIIIKKKQQNEQRIVINAEYDSSLINKDEEI
                     KKEINNQIIELNKHNENISNIFKDIQNIKKQSQDIITNMNDMYKSTILLVDIIQKKEE
                     ALNKQKNILRNIDNILNKKENIIDKVIKCNCDDYKDILIQNETEYQKLQNINHTYEEK
                     KKSIDILKIKNIKQKNIQEYKNKLEQMNTIINQSIEQHVFINADILQNEKIKLEEIIK
                     NLDILDEQIMTYHNSIDELYKLGIQCDNHLITTISVVVNKNTTKIMIHIKKQKEDIQK
                     INNYIQTNYNIINEEALQFHRLYGHNLISEDDKNNLVHIIKEQKNIYTQKEIDISKII
                     KHVKKGLYSLNEHDMNHDTHMNIINEHINNNILQPYTQLINMIKDIDNVFIKIQNNKF
                     EQIQKYIEIIKSLEQLNKNINTDNLNKLKDTQNKLINIETEMKHKQKQLINKMNDIEK
                     DNITDQYMHDVQQNIFEPITLKMNEYNTLLNDNHNNNINNEHQFNHLNSLHTKIFSHN
                     YNKEQQQEYITNIMQRIDVFINDLDTYQYEYYFYEWNQEYKQIDKNKINQHINNIKNN
                     LIHVKKQFEHTLENIKNNENIFDNIQLKKKDIDDIIININNTKETYLKELNKKKNVTK
                     "
ORIGIN

Alignment Scores:
Pred. No.:                1.95e+03        Length:          9022
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       2               Gaps:            0

US-09-858-332G-28 (1-6) x AF533700 (1-9022)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       8633 CATAATCATAATCATAAT 8650

RESULT 368
AF411930
LOCUS       AF411930                9024 bp    DNA     linear   INV 04-AUG-2003
DEFINITION  Plasmodium falciparum strain HB3 normocyte-binding protein 1 (NBP1)
            gene, complete cds.
ACCESSION   AF411930
VERSION     AF411930.2  GI:33414602
KEYWORDS    .
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 9024)
  AUTHORS   Rayner,J.C., Vargas-Serrato,E., Huber,C.S., Galinski,M.R. and
            Barnwell,J.W.
  TITLE     A Plasmodium falciparum Homologue of Plasmodium vivax Reticulocyte
            Binding Protein (PvRBP1) Defines a Trypsin-resistant Erythrocyte
            Invasion Pathway
  JOURNAL   J. Exp. Med. 194 (11), 1571-1582 (2001)
  PUBMED    11733572
REFERENCE   2  (bases 1 to 9024)
  AUTHORS   Rayner,J.C., Barnwell,J.W., Galinski,M.R., Tran,T., Huber,C.S. and
            Corredor,V.
  TITLE     Differing geographic diversity among members of the Reticulocyte
            Binding Like (RBL) gene family of the two major human malaria
            parasites, Plasmodium falciparum and Plasmodium vivax
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 9024)
  AUTHORS   Rayner,J.C., Galinski,M.R. and Barnwell,J.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-AUG-2001) Division of Parasitic Diseases, Centers for
            Disease Control and Prevention, 4770 Buford Highway, MS F-13,
            Chamblee, GA 30341, USA
REFERENCE   4  (bases 1 to 9024)
  AUTHORS   Rayner,J.C., Barnwell,J.W. and Galinski,M.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-2002) Division of Parasitic Diseases, Centers for
            Disease Control and Prevention, 4770 Buford Highway, MS F-13,
            Chamblee, GA 30341, USA
  REMARK    Sequence update by submitter
COMMENT     On Aug 4, 2003 this sequence version replaced gi:17352157.
FEATURES             Location/Qualifiers
     source          1. .9024
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /strain="HB3"
                     /db_xref="taxon:5833"
     gene            <8. .>9024
                     /gene="NBP1"
     mRNA            join(<8. .62,170. .>9024)
                     /gene="NBP1"
                     /product="normocyte-binding protein 1"
     CDS             join(8. .62,170. .9024)
                     /gene="NBP1"
                     /note="similar to Plasmodium vivax reticulocyte binding
```

protein 1"
/codon_start=1
/product="normocyte-binding protein 1"
/protein_id="AAL38219.2"
/db_xref="GI:33414603"
/translation="MQRWIFCNIVLHILIYLAEFSHEQESYSSNEKIRKDYSDDNNYE
PTPSYEKRKKEYGKDESYIKNYRGNNFSYDLSKNSSIFLHMGNGSNSKTLKRCNKKKN
IKTNFLRPIEEEKTVLNNYVYKGVNFLDTIKRNDSSYKFDVYKDTSFLKNREYKELIT
MQYDYAYLEATKEVLYLIPKDKDYHKFYKNKLEKILFNLKDSLKLLREGYIQSKLEMI
RIHSDIDILNEFHQGNIINDNYFNNEIKKKKEDMEKYIREYNLYIYKYENQLKIKIQK
LTNEVSINLNKSTCEKNCYNYILKLEKYKNIIKDKINKWKDLPEIYIDDKSFSYTFLK
DVINNKIDIYKTISSFISTQKQLYYLEYIYIMNKNTLNLLSYNIQKTDINSSSKYTYT
KSHFLKDNHILLSKYYTAKFIDILNKTYYYNLYKNKILLFNKYIIKLRNDLKEYAFKS
IQFIQDKIKKHKDELSIENILQEVNNIYIKYDTSINEISKYNNLIINTDLQIVQQKLL
EIKQKKNDITHKVQLINHIYKNIHDEILNKKNNEITKIIINNIKDHKKDLQDLLLFIQ
QIKQYNILTDHNITQCNNYYKEIIKMKEDINHIHIYIQPILNNLHTLKQVQNNKIKYE
EHIKQILQKIYDKKESLKKIILLKDEAQLDITLLDDLIQKQTKKQTQTQTQTQKQTLI
QNNETIQLISGQEDKHESNPFNHIQTYIQQKDTQNKNIQNLLKSLYNGNINTFIDTIS
KYILKQKDIELTQHVYTDEKINDYLEEIKNEQNKIDKTIDDIKIQETLKQITHIVNNI
KTIKKDLLKEFIQHLIKYMNERYQNMQQGYNNLTNYINQYEEENNNMKQYITTIRNIQ
KIYYDNIYAKEKEIRSGQYYKDFITSRKNIYNIRENISKNVDMIKNEEKKKIQNCVDK
YNSIKQYVKMFKNGDTQDENNNNNNDIYDKLIVPLDSIKQNIDKYNTEHNFITFTNKI
NTHNKKNQEMMEEFIYAYKRLKILKILNISLKACEKNNKSINTLNDKTQELKKIVTHE
IDLLQKDILTSQISNKNVLLLNDLLKEIEQYIIDVHKLKKKSNDLFTYYEQSKNYFYF
KNKKDNFDIQKTINKMNEWLAIKNYINEINKNYQTLYEKKINVLLHNSKSYVQYFYDH
IINLILQKKNYLENTLKTKIQDNEHSLYALQQNEEYQKVKNEKDQNEIKKIKQLIEKN
KNDILTYENNIEQIEQKNIELKTNAQNKDDQIVNTLNEVKKKIIYTYEKVDNQISNVL
KNYEEGKVEYDKNVVQNVNDADDTNDIDEINDIDEINDIDEIKDIDHIKHFDDTKHFD
DIYHADDTRDEYHIALSNYIKTELRNINLQEIKNNIIKIFKEFKSAHKEIKKESEQIN
KEFTKMDVVINQLRDIDRQMLDLYKELDEKYSEFNKTKIEEINNIRENINNVEIWYEK
NIIEYFLRHMNDQKDKAAKYMENIDTYKNNIEIISKQINPENYVETLNKSNMYSYVEK
ANDLFYKQINNIIINSNQLKNEAFTIDELQNIQKNRKNLLTKKQQIIQYTNEIENIFN
EIKNINNILVLTNYKSILQDISQNINHVSIYTEQLHNLYIKLEEEKEQMKTLYHKSNV
LHNQINFNEDAFINNLLINIEKIKNDITHIKEKTNIYMIDVNKSKNNAQLYFHNTLRG
NEKIEYLKNLKNSTNQQITLQELKQVQENVEKVKDIYNQTIKYEEEIKKNYHIITDYE
NKINDILHNSFIKQINMESSNNKKQTKQIIDIINDKTFEEHIKTSKTKINMLKEQSQM
KHIDKTLLNEQALKLFVDINSTNNNLDNMLSEINSIQNNIHTYIQEANKSFDKFKIIC
DQNVNDLLNKLSLGDLNYMNHLKNLQNEIRNMNLEKNFMLDKSKKIDEEEKKLDILKV
NISNINNSLDKLKKYYEEALFQKVKEKAEIQKENIEKIKQEINTLSDVFKKPFFFIQL
NTDSSQHEKDINNNVETYKNNIDEIYNVFIQSYNLIQKYSSEIFSSTLNYIQTKEIKE
KSIKEQNQLNQNEKEASVLLKNIKINETIKLFKQIKNERQNDVHNIKEDYNLLQQYLN
YMKNEMEQLKKYKIDVHMDKNYVENNNGEKELLKETISSYYDKINNINNKLYIYKNK
EDTYFNNMIKVSEILNIIIKKKQQNEQRIVINAEYDSSLINKDEEIKKEINNQIIELN
KHNENISNIFKDIQNIKKQSQDIITNMNDMCKSTILLVDIIQKKEEALNKQKNILRNI
DNILNKKENIIDKVIKCNCDDYKDILIQNETEYQKLQNINHTYEEKKKSIDILKIKNI
KQKNIQEYKNKLEQMNTIINQSIEQHVFINADILQNEKIKLEEIIKNLDILDEQIMTY
HNSIDELYKLGIQCDNHLITTISVVVNKNTTKIMIHIKKQKEDIQKINNYIQTNYNII
NEEALQFHRLYGHNLISEDDKNNLVHIIKEQKNIYTQKEIDISKIIKHVKKGLYSLNE
HDMNHDTHMNIINEHINNNILQPYTQLINMIKDIDNVFIKIQNNKFEQIQKYIEIIKS
LEQLNKNINTDNLNKLKDTQNKLINIETEMKHKQKQLINKMNDIEKDNITDQYMHDVQ
QNIFEPITLKMNEYNTLLNDDHNNNINNEHQFNHLNSLHTKIFSHNYNKEQQQEYITN
IMQRIDVFINDLDTYQYEYYFYEWNQEYKQIDKNKINQHINNIKNNLIHVKKQFEHTL
ENIKNNENIFDNIQLKKKDIDDIIININNTKETYLKELNKKKMLQNKKKVDEKSEINN
HHTLQHDNQNVEQKNKIKDHNLITKPNNNSSEESHQNEQMKEQNKNILEKQTRNIKPH
HVHNHNHNHNHNHNHNHNHNQNQKDSTKLQEQDISTHKLHNTIHEQQSKDNHQGNREK
KQKNGNHERMYFASGIVVSILFLSSLGFVINSKNNKQEYDKEQEKQQQNDFVCDNNKM
DDKSTQKYGRNQEEVMEISFDNDYI"

ORIGIN

Alignment Scores:

Pred. No.:	1.95e+03	Length:	9024
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-28 (1-6) x AF411930 (1-9024)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       8605 CATAATCATAATCATAAT 8622
```

RESULT 369
AF411933
LOCUS AF411933 9044 bp DNA linear INV 04-AUG-2003
DEFINITION Plasmodium falciparum strain 7G8 normocyte-binding protein 1 (NBP1) gene, complete cds.
ACCESSION AF411933
VERSION AF411933.2 GI:33414608
KEYWORDS .
SOURCE Plasmodium falciparum (malaria parasite P. falciparum)
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 9044)
AUTHORS Rayner,J.C., Vargas-Serrato,E., Huber,C.S., Galinski,M.R. and Barnwell,J.W.
TITLE A Plasmodium falciparum Homologue of Plasmodium vivax Reticulocyte Binding Protein (PvRBP1) Defines a Trypsin-resistant Erythrocyte Invasion Pathway
JOURNAL J. Exp. Med. 194 (11), 1571-1582 (2001)
PUBMED 11733572
REFERENCE 2 (bases 1 to 9044)
AUTHORS Rayner,J.C., Barnwell,J.W., Galinski,M.R., Tran,T., Huber,C.S. and Corredor,V.
TITLE Differing geographic diversity among members of the Reticulocyte Binding Like (RBL) gene family of the two major human malaria parasites, Plasmodium falciparum and Plasmodium vivax
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 9044)
AUTHORS Rayner,J.C., Galinski,M.R. and Barnwell,J.W.
TITLE Direct Submission
JOURNAL Submitted (22-AUG-2001) Division of Parasitic Diseases, Centers for Disease Control and Prevention, 4770 Buford Highway, MS F-13, Chamblee, GA 30341, USA
REFERENCE 4 (bases 1 to 9044)
AUTHORS Rayner,J.C., Barnwell,J.W. and Galinski,M.R.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-2002) Division of Parasitic Diseases, Centers for Disease Control and Prevention, 4770 Buford Highway, MS F-13, Chamblee, GA 30341, USA
REMARK Sequence update by submitter
COMMENT On Aug 4, 2003 this sequence version replaced gi:17352163.
FEATURES Location/Qualifiers
source 1. .9044
/organism="Plasmodium falciparum"
/mol_type="genomic DNA"
/strain="7G8"
/db_xref="taxon:5833"
gene <8. .>9026
/gene="NBP1"
mRNA join(<8. .62,151. .>9026)
/gene="NBP1"
/product="normocyte-binding protein 1"
CDS join(8. .62,151. .9026)
/gene="NBP1"
/note="similar to Plasmodium vivax reticulocyte binding protein 1"
/codon_start=1
/product="normocyte-binding protein 1"
/protein_id="AAL38222.2"
/db_xref="GI:33414609"
/translation="MQRWIFCNIVLHILIYLAEFSHEQESYSSNEKIRKDYSDDNNYE
PTPSYEKRKKEYGKDESYIKNYRGNNFSYDLSKNSSIFLHMGNGSNSKTLKRCNKKKN
IKTNFLRPIEEEKTVLNNYVYKGVNFLDTIKRNDSSYKFDVYKDTSFLKNREYKELIT
MQYDYAYLEATKEVLYLIPKDKDYHKFYKNELEEILFNLKDSLKLLREGYIQSKLEMI
RIHSDIDILNEFHQGNIINDNYFNNEIKKKKEDVEKYIREYNLYIYKYENQLKIKIQK
LTNEVSINLNKSTCEKNCYNYILKLEKYKNIIKDKINKWKDLPEIYIDDKSFSYTFLK
DVINNKIDIYKTISSFISTQKQLYYLEYIYIMNKNTLNLLSYNIQKTDINSSSKYTYT
KSHFLKDNHILLSKYYTAKFIDILNKTYYYNLYKNKILLFNKYIINLRNDLKEYAFKS
IQFIQDKIKKHKDELSIENILQEVNNIYIKYDTSINEISKYNNLIINTDLQIVQQKLL
EIKQKKNDITHKVQLINHIYKNIHDEILNKKNNEITKIIINNIKDHKKDLQDLLLFIQ
QIKQYNILTDHNITQCNNYYKEIIKMKEDINHIHIYIQPILNNLHTLKQVQNNKIKYE
EHIKQILQKIYDKKESLKKIILLKDEAQLDITLLDDLIQKQTKKQTQTQTQTQKQTLI
QNNETIQLISGQEDKHESNPFNHIQTYIQQKDTQNKNIQNLLKSLYNGNINTFIDTIS
KYILKQKDIELTQHVYTDEKINDYLEEIKNEQNKIDKTIDDIKIQETLKQITHIVNNI

KTIKKDLLKEFIQHLIKYMNERYQNMQQGYNNLTNYINQYEEENNNMKQYITTIRNIQ
KIYYDNIYAKEKEIRSGQYYKDFITSRKNIYNIRENISKNVDMIKNEEKKKIQNCVDK
YNSIKQCVKMLKNGDTQDENNNNNNNNNNNDIYDKLIVPLDSIKQNIDKYNTEHNFIT
FTNKINTHNKKNQEMMEEFIYAYKRLKILKILNISLKACEKNNKSINTLNDKTQELKK
IVTHEIDLLQKDILTSQISNKNVLLLNDLLKEIEQYIIDVHKLKKKSNDLFTYYEQSK
NYFYFKNKKDNFDIQKTINKMNEWLAIKNYINEINKNYQTLYEKKINVLLHNSKSYVQ
YFYDHIINLILQKKNYLENTLKTKIQDNEHSLYALQQNEEYQKVKNEKDQNEIKKIKQ
LIEKNKNDILTYENNIEQIEQKNIELKTNAQNKDDQIVNTLNEVKKKIIYTYEKVDNQ
ISNVLKNYEEGKVEYDKNVVQNVNDADDTNDIDEINDIDEINDIDEINDIDEIKDIDH
IKHFDDTKHFDDIYHADDTRDEYHIALSNYIKTELRNINLQEIKNNIIKIFKEFKSAH
KEIKKESKQINKEFTKMDVVINQLRDIDRQMLDLYKELDEKYSEFNKTKIEEINNIRE
NINNVEIWYEKNIIEYFLRHMNDQKDKAAKYMENIDTYKNNIEIISKQINPENYVETL
NKSNMYSYVEKANDLFYKQINNIIINSNQLKNEAFTIDELQNIQKNRKNLLTKKQQII
QYTNEIENIFNEIKNINNILVLTNYKSILQDISQNINHVSIYTEQLHNLYIKLEEEKE
QMKTLYHKSNVLHNQINFNEDAFINNLLINIEKIKNDITHIKEKTNIYMIDVNKSKNN
AQLYFHNTLRGNEKIEYLKNLKNSTNQQITLQELKQVQENVEKVKDIYNQTIKYEEKI
KKNYHIITDYENKINDILHNSFIKQINMESSNNKKQTKQIIDIINDKTFEEHIKTSKT
KINMLKEQSQMKHIDKTLLNEQALKLFVDINSTNNNLDNMLSEINSIQNNIHTYIQEA
NKSFDKFKIICDQNVNDLLNKLSLGDLNYMNHLKNLQNEIRNMNLEKNFMLDKSKKID
EEEKKLDILKVNISNINNSLDKLKKYYEEALFQKVKEKAENQKENIEKIKQEINTLSD
VFKKPFFFIQLNTDSSQHEKDINNNVETYKNNIDEIYNVFIQSYNLIQKYSSEIFSST
LNYIQTKEIKEKSIKEQNQLNQNEKEASVLLKNIKINETIKLFKQIKNERQNDVHNIK
EDYNLLQQYLNYMKNEMEQLKKYKNDVHMDKNYVENNNGEKEKLLKETISSYYDKINN
INNKLYIYKNKEDTYFNNMIKVSEILNIIIKKKQQNEQRIVINAEYDSSLINKDEEIK
KEINNQIIELNKHNENISNIFKDIQNIKKQSQDIITNMNDMHKSTILLVDIIQKKEEA
LNKQKNILRNIDNILNKKENIIDKVIKCNCDDYKDILIQNETEYQKLQNINHTYEEKK
KSIDILKIKNIKQKNIQEYKNKLEQMNTIINQSIEQHVFINADILQNEKIKLEEIIKN
LDILDEQIMTYHNSIDELYKLGIQCDNHLITTISVVVNKNTTKIMIHIKKQKEDIQKI
NNYIQTNYNIINEEALQFHRLYGHNLISEDDKNNLVHIIKEQKNIYTQKEIDISKIIK
HVKKGLYSLNEHDMNHDTHMNIINEHINNNILQPYTQLINMIKDIDNVFIKIQNNKFE
QIQKYIEIIKSLEQLNKNINTDNLNKLKDTQNKLINIETEMKHKQKQLINKMNDIEKD
NITDQYMHDVQQNIFEPITLKMNEYNTLLNDNHNNNINNEHQFNHLNSLHTKIFSHNY
NKEQQQEYITNIMQRIDVFINDLDTYQYEYYFYEWNQEYKQIDKNKINQHINNIKNNL
IHVKKQFEHTLENIKNNENIFDNIQLKKKDIDDIIINNNTKETYLKELNKKKMLQNK
KKVDEKSEINNHHTLQHDNQNVEQKNKIKDHNLITKPNNNSSEESHQNEQMKEQNKNI
LEKQTRNIKPHHVHNHNHNHNHNHNQNQNQKDSTKLQEQDISTHKLHNTIHEQQSKDN
HQGNREKKQKNGNHERMYFASGIVVSILFLSSLGFVINSKNNKQEYDKEQEKQQQNDF
VCDNNKMDDKSTQKYGRNQEEVMEISFDNDYI"

ORIGIN

Alignment Scores:

Pred. No.:	1.95e+03	Length:	9044
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-28 (1-6) x AF411933 (1-9044)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       8619 CATAATCATAATCATAAT 8636
```

RESULT 370
AF411929
LOCUS AF411929 9049 bp DNA linear INV 04-AUG-2003
DEFINITION Plasmodium falciparum strain FVO normocyte-binding protein 1 (NBP1) gene, complete cds.
ACCESSION AF411929
VERSION AF411929.2 GI:33414600
KEYWORDS .
SOURCE Plasmodium falciparum (malaria parasite P. falciparum)
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 9049)
AUTHORS Rayner,J.C., Vargas-Serrato,E., Huber,C.S., Galinski,M.R. and Barnwell,J.W.
TITLE A Plasmodium falciparum Homologue of Plasmodium vivax Reticulocyte Binding Protein (PvRBP1) Defines a Trypsin-resistant Erythrocyte Invasion Pathway
JOURNAL J. Exp. Med. 194 (11), 1571-1582 (2001)
PUBMED 11733572
REFERENCE 2 (bases 1 to 9049)
AUTHORS Rayner,J.C., Barnwell,J.W., Galinski,M.R., Tran,T., Huber,C.S. and Corredor,V.
TITLE Differing geographic diversity among members of the Reticulocyte Binding Like (RBL) gene family of the two major human malaria parasites, Plasmodium falciparum and Plasmodium vivax
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 9049)
AUTHORS Rayner,J.C., Barnwell,J.W. and Galinski,M.R.
TITLE Direct Submission
JOURNAL Submitted (22-AUG-2001) Division of Parasitic Diseases, Centers for Disease Control and Prevention, 4770 Buford Highway, MS F-13, Chamblee, GA 30341, USA
REFERENCE 4 (bases 1 to 9049)
AUTHORS Rayner,J.C., Barnwell,J.W. and Galinski,M.R.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-2002) Division of Parasitic Diseases, Centers for Disease Control and Prevention, 4770 Buford Highway, MS F-13, Chamblee, GA 30341, USA
REMARK Sequence update by submitter
COMMENT On Aug 4, 2003 this sequence version replaced gi:17352155.
FEATURES Location/Qualifiers
source 1. .9049
/organism="Plasmodium falciparum"
/mol_type="genomic DNA"
/strain="FVO"
/db_xref="taxon:5833"
gene <8. .>9048
/gene="NBP1"
mRNA join(<8. .62,170. .>9048)
/gene="NBP1"
/product="normocyte-binding protein 1"
CDS join(8. .62,170. .9048)
/gene="NBP1"
/note="similar to Plasmodium vivax reticulocyte binding protein 1"
/codon_start=1
/product="normocyte-binding protein 1"
/protein_id="AAL38218.2"
/db_xref="GI:33414601"
/translation="MQRWIFCNIVLHILIYLAEFSHEQESYSSNEKIRKDYSDDNNYE
PTPSYEKRKKEYGKDESYIKNYRGNNFSYDLSKNSSIFLHMGNGSNSKTLKRCNKKKN
IKTNFLRPIEEEKTVLNNYVYKGVNFLDTIKRNDSSYKFDVYKDTSFLKNREYKELIT
MQYDYAYLEATKEVLYLIPKDKDYHKFYKNKLEKILFNLKDSLKLLREGYIQSKLEMI
RIHSDIDILNEFHQGNIINDNYFNNEIKKKKEDMEKYIREYNLYIYKYENQLKIEIQK
LTNEVSINLNKSTCEKNCYYYILKLEKYKNIIKDKINKWKDLPEIYIDDKSFSYTFLK
DVINNKIDIYKTISSFISTQKQLYYLEYIYIMNKNTLNLLSYNIQKTDINSSSKYTYT
KSHFLKDNHILLSKYYTAKFIDILNKTYYYNLYKNKILLFNKYIIKLRNDLKEYAFKS
IQFIQDKIKKHKDELSIENILQEVNNIYIKYDTSINEISKYNNLIINTDLQIVQQKLL
EIKQKKNDITHKVQLINHIYKNIHDEILNKKNNEITKIIINNIKDHKKDLQDLLLFIQ
QIKQYNILTDHNITQCNNYYKEIIKMKEDINHIHIYIQPILNNLHTLKQVQNNKIKYE
EHIKQILQKIYDKKESLKKIILLKDEAQLDITLLDDLIQKQTQTQTQTQKQTLIQNNE
TIQLISGQEDKHESNPFNHIQTYIQQKDTQNKNIQNLLKSLYNGNINTFIDTISKYIL
KQKDIELTQHVYTDEKINDYLEEIKNEQNKIDKTIDDIKIQETLKQITHIVNNIKTIK
KDLLKEFIQHLIKYMNERYQNMQQGYNNLTNYINQYEEENNNMKQYITTIRNIQKIYY
DNIYAKEKEIRSGQYYKDFITSRKNIYNIRENISKNVDMIKNEEKKKIQNCVDKYNSI
KQYVKMFKNGDTQDENNNNNNDIYDKLIVPLDSIKQNIDKYNTEHNFITFTNKINTHN
KKNQEMMEEFIYVYKRLKILKILNISLKACEKNNKSINTLNDKTQELKKIVTHEIDLL
QKDILTSQISNKNVLLLNDLLKEIEQYIIDVHKLKKKSNDLFTYYEQSKNYFYFKNKK
DNFDIQKTINKMNEWLAIKNYINEINKNYQTLYEKKINVLLHNSKSYVQYFYDHIINL
ILQKKNYLENTLKTKIQDNEHSLYALQQNEEYQKVKNEKDQNEIKKIKQLIEKNKNDI
LTYENNIEQIEQKNIELKTNAQNKDDQIVNTLNEVKKKIIYTYEKVDNQISNVLKNYE
EGKVEYDKNVVQNVNDADDTNDIDEINDIDEINDIDEINDIDEINDIDEIKDIDHIKH
FDDTKHFDDIYHADDTRDEYHIALSNYIKTELRNINLQEIKNNIIKIFKEFESAHKEI
KKESKQINKEFTKMDVVINQLRDIDRQMLDLYKELDEKYSEFNKTKIEEINNIRENIN
NVEIWYEKNIIEYFLRHMNDQKDKAAKYMENIDTYKNNIEIISKQINPENYVETLNKS
NMYSYVEKANDLFYKQINNIIINSNQLKNEAFTIDELQNIQKNRKNLLTKKQQIIQYT
NEIENIFNEIKNINNILVLTNYKSILQDISQNINHVSIYTEQLHNLYIKLEEEKEQMK
TLYHKSNVLHNQINFNEDAFINNLLINIEKIKNDITHIKEKTNIYMIDVNKSKNNAQL
YFHNTLRGNEKIEYLKNLKNSTNQQITLQELKQVQENVEKVKDIYNQTIKYEEKIKKN
YHIITDYENKINDILHNSFIKQINMESSNNKKQTKQIIDIINDKTFEEHIKTSKTKIN
MLKEQSQMKHIDKTLLNEQALKLFVDINSTNNNLDNMLSEINSIQNNIHTYIQEANKS
FDKFKIICDQNVNDLLNKLSLGDLNYMNHLKNLQNEIRNMNLEKNFMLDKSKKIDEEE

```
                     KKLDILKVNISNINNSLDKLKKYYEEALFQKVKEKAENQKENIEKIKQEINTLSDVFK
                     KPFFFIQLNTDSSQHEKDINNNVETYKNNIDEIYNVFIQSYNLIQKYSSEIFSSTLNY
                     IQTKEIKEKSIKEQNQLNQNEKEAFVLLKNIKINETIKLFKQIKNERQNDVHNIKEDY
                     NLLQQYLNYMKNEMEQLKKYKNDVHMDKNYVENNNGEKEKLLKETISSYYDKINNINN
                     KLYIYKNKEDTYFNNMIKVSEILNIIIKKKQQNEQRIVINAEYDSSLINKDEEIKKEI
                     NNQIIELNKHNENISNIPKDIQNIKKQSQDIITNMNDMHKSTILLVDIIQKKEEALNK
                     QKNILRNIDNILNKKENIIDKVIKCNCDDYKDILIQNETEYQKLQNINHTYEEKKKSI
                     DILKIKNIKQKNIQEYKNKLEQMNTIINQSIEQHVFINADILQNEKIKLEEIIKNLDI
                     LDEQIMTYHNSIDELYKLGIQCDNHLITTISVVVNKNTTKIMIHIKKQKEDIQKINNY
                     IQTNYNIINEEALQFHRLYGHNLISEDDKNNLVHIIKEQKNIYTQKEIDISKIIKHVK
                     KGLYSLNEHDMNHDTHMNIINEHINNNILQPYTQLINMIKDIDNVFIKIQNNKFEQIQ
                     KYIEIIKSLEQLNKNINTDNLNKLKDTQNKLINIETEMKHKQKQLINKMNDIEKDNIT
                     DQYMHDVQQNIFEPITLKMNEYNTLLNDDHNNNINNEHQFNHLNSLHTKIFSHNYNKE
                     QQQEYITNIMQRIDVFINDLDTYQYEYYFYEWNQEYKQIDKNKINQHINNIKNNLIHV
                     KKQFEHTLENIKNNENIFDNIQLKKKDIDDIIININNTKETYLKELNKKKMLQNKKKV
                     DEKSEINNHHTLQHDNQNVEQKNKIKDHNLITKPNNNSSEESHQNEQMKEQNKNILEK
                     QTRNIKPHHVHNHNHNHNHNHNHNHNQNQKDSTKLQEQDISTHKLHNTIHEQQSKD
                     NHQGNREKKQKNGNHERMYFASGIVVSILFLSSLGFVINSKNNKQEYDKEQEKQQQND
                     FVCDNNKMDDKSTQKYGRNQEEVMEISFDNDYI"
ORIGIN

Alignment Scores:
Pred. No.:              1.95e+03        Length:         9049
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-858-332G-28 (1-6) x AF411929 (1-9049)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       8629 CATAATCATAATCATAAT 8646

RESULT 371
BMCHORIO/c
LOCUS       BMCHORIO                9454 bp    DNA     linear   INV 15-FEB-1999
DEFINITION  Bombyx mori chorion locus sequence encompassing gene pairs A/B.L12
            and A/B.L11, pseudogene PSI-A.L11 and retrotransposon BM-A.L11.
ACCESSION   X15557 X04028 X04029 X04030 X04031 X04032 X04033 X04034 X04035
VERSION     X15557.1  GI:5820
KEYWORDS    chorion protein; pseudogene; retrotransposon.
SOURCE      Bombyx mori (domestic silkworm)
  ORGANISM  Bombyx mori
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
            Bombycoidea; Bombycidae; Bombyx.
REFERENCE   1
  AUTHORS   Spoerel,N., Nguyen,H.T. and Kafatos,F.C.
  TITLE     Gene regulation and evolution in the chorion locus of Bombyx mori.
            Structural and developmental characterization of four eggshell
            genes and their flanking DNA regions
  JOURNAL   J. Mol. Biol. 190 (1), 23-35 (1986)
   PUBMED   3023635
REFERENCE   2
  AUTHORS   Mitsialis,S.A., Spoerel,N., Leviten,M. and Kafatos,F.C.
  TITLE     A short 5'-flanking DNA region is sufficient for developmentally
            correct expression of moth chorion genes in Drosophila
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 84 (22), 7987-7991 (1987)
   PUBMED   3120180
REFERENCE   3  (bases 1 to 9454)
  AUTHORS   Spoerel,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUN-1989) Spoerel N., Dept of Biochemistry,
            University of Conneticut, Health Centre, Parmington CT 06032, U S A
FEATURES             Location/Qualifiers
     source          1. .9454
                     /organism="Bombyx mori"
                     /mol_type="genomic DNA"
                     /strain="703"
                     /db_xref="taxon:7091"
                     /clone="B2"
     gene            complement(464. .1026)
                     /gene="A.L12"
     mRNA            complement(join(464. .866,976. .1026))
                     /gene="A.L12"
     exon            complement(464. .866)
                     /gene="A.L12"
                     /number=2
     old_sequence    491
                     /citation=[1]
                     /replace="c"
     CDS             complement(join(519. .866,976. .1026))
                     /gene="A.L12"
                     /codon_start=1
                     /protein_id="CAA33563.1"
                     /db_xref="GI:5821"
                     /db_xref="GOA:P08825"
                     /db_xref="UniProt/Swiss-Prot:P08825"
                     /translation="MSTFAFLLFCAQACLIQNVCGLYLGGLGGCGCGGGWNGWNGLGG
                     GWNGLGVGWNGIDGGYGAGCGSYGGEGIGNVGVAGELPVGGVTAVGGRVPIIGGVEYG
                     GPACAAGAVSICGHCAPTCGCGRAGLGGYY"
     intron          complement(867. .975)
                     /gene="A.L12"
                     /number=1
     exon            complement(976. .1056)
                     /number=1
     TATA_signal     complement(1081. .1087)
                     /note="TATA-Box of A.L12"
     misc_signal     1266. .1271
                     /note="putative transcription signal"
     TATA_signal     1303. .1309
                     /note="TATA-Box of B.L12"
     gene            1333. .1995
                     /gene="B.L12"
     mRNA            join(1333. .1409,1493. .1995)
                     /gene="B.L12"
     exon            1333. .1409
                     /gene="B.L12"
                     /number=1
     CDS             join(1359. .1409,1493. .1927)
                     /gene="B.L12"
                     /codon_start=1
                     /protein_id="CAA33564.1"
                     /db_xref="GI:5822"
                     /db_xref="GOA:P08828"
                     /db_xref="UniProt/Swiss-Prot:P08828"
                     /translation="MAAKLILFVCATALVAQSVLSIGCGCGGRGYGGLGYGGLGYGGL
                     GGGCGRGFSGGGLPVATASAAPTGLGIASENRYEGTVGVCGNLPFLGTADVAGEFPTA
                     GIGEIDYGCGNGAVGITREGGFGYGAGYGDGYGLGFGGYGGGYGLGNGGYGGCGCGWG
                     Y"
     intron          1410. .1492
                     /gene="B.L12"
                     /number=1
     exon            1493. .1995
                     /gene="B.L12"
                     /number=2
     misc_feature    complement(3972. .4103)
                     /note="PSI-A.L11 pseudogene"
                     /pseudo
     gene            complement(4379. .7165)
                     /gene="A.L11"
     mRNA            complement(join(4379. .4813,7083. .7165))
                     /gene="A.L11"
     exon            complement(4379. .4813)
                     /gene="A.L11"
                     /number=2
     CDS             complement(join(4475. .4813,7083. .7133))
                     /gene="A.L11"
                     /codon_start=1
                     /protein_id="CAA33565.1"
                     /db_xref="GI:4376181"
                     /db_xref="GOA:P08826"
                     /db_xref="UniProt/Swiss-Prot:P08826"
                     /translation="MSTFAFLLLCVQACLIQNVYGQCLGRGLGGCGCGGGLGGYGLGY
```

```
                     GLGGYGGGYGYGGYGGGYYGGYGGEGVGNVGVAGVLPVGGVTAVGGRVPIIGGVEFGG
                     PACAGGCVSICGHCAPTCGCGYGGLYY"
     intron          complement(4814. .7082)
                     /gene="A.L11"
                     /number=1
     repeat_unit     4916. .4919
                     /note="insertion target site"
     repeat_unit     4920. .5054
                     /note="LTR of transposon BM- A.L11"
                     /rpt_type=TERMINAL
     repeat_unit     6910. .7044
                     /note="LTR of transposon BM- A.L11"
                     /rpt_type=TERMINAL
     repeat_unit     7045. .7048
                     /note="insertion target site"
     exon            complement(7083. .7165)
                     /gene="A.L11"
                     /number=1
     TATA_signal     complement(7190. .7196)
                     /note="TATA-Box of A.L11"
     misc_signal     7356. .7361
                     /note="putative transcription signal"
     TATA_signal     7413. .7419
                     /note="TATA-Box of B.L11"
     gene            7443. .8462
                     /gene="B.L11"
     mRNA            join(7443. .7519,7939. .8462)
                     /gene="B.L11"
     exon            7443. .7519
                     /gene="B.L11"
                     /number=1
     CDS             join(7469. .7519,7939. .8394)
                     /gene="B.L11"
                     /codon_start=1
                     /protein_id="CAA33566.1"
                     /db_xref="GI:5824"
                     /db_xref="GOA:P08827"
                     /db_xref="UniProt/Swiss-Prot:P08827"
                     /translation="MAAKLILFVCATALVAQSALGIGCGCGCGGRGYGGLGYGGLGYG
                     GLGYGGYGGGCGRGFSGGGLPVATASAAPTGLGIASENRYEGTVGVCGNLPFLGTADV
                     AGEFPTAGIGEILYGCGNGAVGITREGGLGFGAGYGGGYGLGYGGYGGGYGLGYGGYG
                     GCGCGCGY"
     intron          7520. .7938
                     /gene="B.L11"
                     /number=1
     exon            7939. .8462
                     /gene="B.L11"
                     /number=2
ORIGIN

Alignment Scores:
Pred. No.:                2.04e+03        Length:          9454
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       2               Gaps:            0

US-09-858-332G-28 (1-6) x BMCHORIO (1-9454)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       7972 CACAACCACAACCACAAC 7955

RESULT 372
CMO580407
LOCUS       CMO580407               9679 bp    RNA     linear   VRL 16-JUN-2005
DEFINITION  Simian immunodeficiency virus SIVden complete proviral genome.
ACCESSION   AJ580407
VERSION     AJ580407.1  GI:39930157
KEYWORDS    env gene; envelope glycoprotein precursor gp160; gag gene; gag
            protein; nef gene; nef protein; pol gene; pol protein; rev gene;
            rev protein; tat gene; tat protein; vif gene; vif protein; vpr
            gene; vpr protein; vpu gene; vpu protein.
SOURCE      Simian immunodeficiency virus
  ORGANISM  Simian immunodeficiency virus
            Viruses; Retro-transcribing viruses; Retroviridae;
            Orthoretrovirinae; Lentivirus; Primate lentivirus group.
REFERENCE   1
  AUTHORS   Dazza,M.C., Ekwalanga,M., Nende,M., Shamamba,K.B., Bitshi,P.,
            Paraskevis,D. and Saragosti,S.
  TITLE     Characterization of a Novel vpu-Harboring Simian Immunodeficiency
            Virus from a Dent's Mona Monkey (Cercopithecus mona denti)
  JOURNAL   J. Virol. 79 (13), 8560-8571 (2005)
  PUBMED    15956597
REFERENCE   2  (bases 1 to 9679)
  AUTHORS   Saragosti,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-AUG-2003) Saragosti S., IMEA INSERM Hopital Bichat,
            IMSERM U552, 46 Rue Henri Huchard, Paris 750189, FRANCE
FEATURES             Location/Qualifiers
     source          1. .9679
                     /organism="Simian immunodeficiency virus"
                     /proviral
                     /mol_type="genomic RNA"
                     /specific_host="Cercopithecus mona subsp. denti"
                     /db_xref="taxon:11723"
     LTR             1. .217
     repeat_region   1. .82
                     /note="R repeat 5' copy"
     repeat_region   83. .217
                     /note="U5 region"
     gene            479. .2155
                     /gene="gag"
     CDS             479. .2155
                     /gene="gag"
                     /codon_start=1
                     /product="gag protein"
                     /protein_id="CAE46398.1"
                     /db_xref="GI:39930158"
                     /db_xref="GOA:Q70IH3"
                     /db_xref="InterPro:IPR000071"
                     /db_xref="InterPro:IPR000721"
                     /db_xref="InterPro:IPR001878"
                     /db_xref="UniProt/TrEMBL:Q70IH3"
                     /translation="MGSGSSVPMTGAALDRFEKVRLRPGSKKKYMIKHLIWSKKEMER
                     FGLGEQLLETYEGCQTIIETLAPLEPQGSESLKSLFRTCCVVWCCHRGLEVADTEEAL
                     REVRKRIGAKKKNIVIGNADQGDDQPPQGAAGGGSSAEQGASAGGSSFRDPNSQNYPI
                     VRTAQGQYQHQALSTRVLKTWIKEVEDRKFSPEIVPLFSVLAEGLIPYDINQLLNAIG
                     EHQGAMQVLKDVINEEAHDWDLAHPANLQQPLPPAAQRDPRGSDIAGTTSSIEDQITW
                     MTRQQGPINVGDIYKRWVIMGLQKVVKMYQPVSILDIKQGPKESFKDYVDRFYKALRA
                     EQADQSVKLWMTTSLLVQNANPDCKQILEALGPNPTLEEMLQACQGVGGPGQKARPTA
                     EALASAFQKQGLCHGSRKGAQPPVERRGLRCFNCGQIGHLAKDCRKPKRVTQPGACFK
                     CGKLGHQAKNCRSNTNDQGQVNFLGNGLGRKNPPRNFVQYYPQMNQDLPWSPVRVQQQ
                     PQYPQNFQEGGNRQSCQEDKRQNKQSHQEDEETTRESLYPSLKSLFGDDQ"
     gene            1975. .4989
                     /gene="pol"
     CDS             1975. .4989
                     /gene="pol"
                     /codon_start=1
                     /product="pol protein"
                     /protein_id="CAE46399.1"
                     /db_xref="GI:39930159"
                     /db_xref="GOA:Q70IH2"
                     /db_xref="HSSP:1B9D"
                     /db_xref="InterPro:IPR000477"
                     /db_xref="InterPro:IPR001037"
                     /db_xref="InterPro:IPR001584"
                     /db_xref="InterPro:IPR001969"
                     /db_xref="InterPro:IPR001995"
                     /db_xref="InterPro:IPR002156"
                     /db_xref="InterPro:IPR003308"
                     /db_xref="InterPro:IPR010659"
                     /db_xref="InterPro:IPR010661"
                     /db_xref="UniProt/TrEMBL:Q70IH2"
                     /translation="MEPSEGPAATTVPTKFSRGGKQTILSGRQEAEQAVPSGGRGDNE
```

```
                GESLSFVEVSLWRRPIGQIWVQGQRVEVLIDTGADDTIIAEQDINLGTEGWIPKTVGG
                IGGYINVQAYPGIEIKLMDKVAVDTVLVGPTPINILGRNYLSKLGVTLNMAVQQMRPQ
                VTPVKLKPGFDGPRIKQWPLTTEKLKALQQIVTQMEKEGRLEKAPPTNPYNTPVFVIP
                KKDKTQYRMLIDFRKLNEATQDFWEVQLGIPHPAGLQQKEHITVLDLKDAYYSVPLDP
                DFAQYTAFTVPSINNEGPGQRYIFKVLPQGWVGSPTIFQATADKMLSQFRKAHPDVIL
                IQYMDDLLVASDRSLEQHRKMVKELRDFLAVWGFETPDAKFQETPPIQWMGYELHPKH
                WQIQKIDLPLPEDGQWTVNAIQKLVGVLNWASQLYSGIKTKELCKSIRGVKDLGETIS
                LTEGAEAELKEAQEILKESISGSYYRPEKPLIAEISKLGENQWGYTIKQEKHLLKTGK
                YHKDKGSHYNPYQQLAKVMAKIGREALVIWGRLPTFRLPVSKVDWDAWWSEYWQVNWI
                PDWEMVSTPPLLRNFYNLVSEPIDKAPTYYTDGAANRQTKQGRAGYVATNGKQKVVTL
                EETTNQKAELIGVLMALQDSPNKVNIVTDSQYALGIIAGTPTTSEHQIVEQIIQAMQG
                KEEIYITWVPAHKGIGGNTEIDKLVSKGIRQILFLEQIPQAQEEHDKYHSNWKEIRDK
                YGLPTMVAKEIINVCPKCQTHGEPIRGQVDASLGTWQMDCTHLENKVIIVAVNAASGY
                LETKILPAETGKETALFLLQIAARWPITKLHTDNGPNFISQAMQAACWWANIKHTTGV
                PYNPQSQGIVENKNKQLKEVIALVREDAERLETALAMATLILNFKRRGGLGDMTAAER
                YINMLNTHLEIQHTKQPKFSNFKVYYREGANPQWQGPAELLWKGEGAVVIKTQAGDIF
                PVPRRKAKIIKPYNAESRKGEQNKGMDSETDI"
gene            4931. .5707
                /gene="vif"
CDS             4931. .5707
                /gene="vif"
                /codon_start=1
                /product="vif protein"
                /protein_id="CAE46400.1"
                /db_xref="GI:39930160"
                /db_xref="GOA:Q70IH4"
                /db_xref="InterPro:IPR000475"
                /db_xref="UniProt/TrEMBL:Q70IH4"
                /translation="MQNLGRENRTKEWIVRLTYKIGETKLEHWMSTLKSIIYKTNELK
                GTVIIPHFRITWEWYTMTKIHIPLRAAYMEVKLYWHLTPEKGYLSQVAISIAYVHLST
                GWATEVDPVTADEIIHKDHFSCFTRQAVRQALRGEVVLSTCARKKAHYHPRTGPGREP
                PSLQYIVLQAAIRYGKVTTITPTSISFKNGRNHSKAATGSSMGYHRGGKKTLQQRRDH
                WNMGSLLVITSAPTLDRRTSHGSRRHRFHKNHAERNMEAL"
gene            5451. .5792
                /gene="vpr"
CDS             5451. .5792
                /gene="vpr"
                /codon_start=1
                /product="vpr protein"
                /protein_id="CAE46401.1"
                /db_xref="GI:39930161"
                /db_xref="GOA:Q70IH1"
                /db_xref="InterPro:IPR000012"
                /db_xref="UniProt/TrEMBL:Q70IH1"
                /translation="MERLPPSHPPALVSRMAETTQRQLQEAVWDITEEARKHFSKEEI
                TGIWDHCWSLPALPHWTEGQVMAAAVIDFIRIMQKEIWKHYRVGCFHREAERVRHYPN
                IRPLRPRREEP"
gene            join(5789. .5974,8440. .8544)
                /gene="tat"
CDS             join(5789. .5974,8440. .8544)
                /gene="tat"
                /codon_start=1
                /product="tat protein"
                /protein_id="CAE46402.1"
                /db_xref="GI:39930162"
                /db_xref="GOA:Q70IH0"
                /db_xref="InterPro:IPR001831"
                /db_xref="UniProt/TrEMBL:Q70IH0"
                /translation="MNADSIDPFAGNKTPCNTCYCKHCAYHCQLCFLQKGLGIVYGRA
                PRKRPRARQSGTPDQEAVSLSDPRRQQISQKKEKEKVETATGSSGGNIREDS"
gene            join(5910. .5974,8440. .8620)
                /gene="rev"
CDS             join(5910. .5974,8440. .8620)
                /gene="rev"
                /codon_start=1
                /product="rev protein"
                /protein_id="CAE46403.1"
                /db_xref="GI:39930163"
                /db_xref="GOA:Q70IG9"
                /db_xref="InterPro:IPR000625"
                /db_xref="UniProt/TrEMBL:Q70IG9"
                /translation="MAGLQERDQELVKAVRLIKKLYPYPTPEGSKSARRRRRRKWRRQ
                QAQVVAISERILEYHLGRSLTAADVELPDLEKLSVSD"
gene            6092. .6286
                /gene="vpu"
CDS             6092. .6286
                /gene="vpu"
                /codon_start=1
                /product="viral protein U"
                /protein_id="CAE46404.1"
                /db_xref="GI:39930164"
                /db_xref="UniProt/TrEMBL:Q70IG8"
                /translation="MLKIKLGPIEYVCLVFAVVITWIAAIGVGYLAYRAYKSYREELR
                YIRLRLWSIDSGYESSQEDP"
gene            6261. .8855
                /gene="env"
CDS             6261. .8855
                /gene="env"
                /codon_start=1
                /product="envelope glycoprotein precursor gp160"
                /protein_id="CAE46405.1"
                /db_xref="GI:39930165"
                /db_xref="GOA:Q70IG7"
                /db_xref="InterPro:IPR000328"
                /db_xref="InterPro:IPR000777"
                /db_xref="UniProt/TrEMBL:Q70IG7"
                /translation="MSQVRRTLKVLLLILWTIGIICTTEKQYVTVFYGIPAWRNSAAP
                LFCTAMDRGVWAAASCVPEDPYQREVWMENATDYFDIYHNPIVDQVVQNMKDLYKQTF
                KPCVKLTPVCVRMNCTRKEDNSTNAVTVSSTTTTTTTIKTPTTEWYGGNKTTIFLKCN
                FNMTTGKGFRDKRVMYEAHCRIEDVMEDRETNQTYYYMRYCNQSTITQACEKVVFEPF
                KIRMCAPAGFALLKCNDHPWSGYGLCKNMSARVCTDEIHTMASTWLIMNSTVELLDNH
                TQIIRNTSKDNQTIAILFRKSSVIKLKCTRPGNVTLKGLNMAAGIVMRATIIPRRAIR
                RAACKIVGGNWTRTRQEIEKAVKEEFNINGTVKWDVPRVGGDPETSSARFQCRGEFFY
                CNLTRMFNSSQYPVNNSNIINNQTYMVCNIKQLVNTWHTVGMDIYLPPAEGEITCTSN
                VSGIIVDTDVHPNRTHITLTLSADVRQVWRAELARWKLVELTPIGLAPTSVRRYEGPE
                SAESVSRRRREVSLVLGLIGFLSAAGTVMGAVGTGLAVQSRDLLAGIVQQQQELLRAV
                EGHSALLQLSVWGIKNLNARLTAIEKYLKDQALLNEWGCAWKQICHTSVEWSNQSIPG
                LALYQPDFQNMTWQQWETDIAALFGNVSDALKKAQAQQERNVHEVQSLYDWDNLWNWF
                DLSKWFWWIRLVVYVIAGLVILRIIMFIINVLSRLCRGYSPLLQIPIRPQKAANQPEE
                GEGESGDGNRLKWWQYPRGFLSIIWEDLSQLLTWSYLTLRNLVFLIRTACSRIWDFVH
                RHSLILLGYFQYGLQELQTGLRDLGTSAIQQGRATAEVVLAALTRAAREVVAIPRRIR
                QGLEIVLN"
gene            8689. .9435
                /gene="nef"
CDS             8689. .9435
                /gene="nef"
                /codon_start=1
                /product="nef protein"
                /protein_id="CAE46406.1"
                /db_xref="GI:39930166"

Alignment Scores:
Pred. No.:              2.08e+03        Length:         9679
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     13              Gaps:           0

US-09-858-332G-28 (1-6) x CMO580407 (1-9679)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       6659 CACAACCACAACCACAAC 6676

RESULT 373
CQ596786
LOCUS       CQ596786              10205 bp    DNA     linear   PAT 02-FEB-2004
DEFINITION  Sequence 24544 from Patent WO0171042.
ACCESSION   CQ596786
VERSION     CQ596786.1  GI:41653170
KEYWORDS    .
SOURCE      Drosophila sp.
  ORGANISM  Drosophila sp.
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
```

```
           Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1
  AUTHORS  Venter,J.C., Adams,M., Li,P.W. and Myers,E.W.
  TITLE    Detection kits, such as nucleic acid arrays, for detecting the
           expression of 10,000 or more Drosophila genes and uses thereof
  JOURNAL  Patent: WO 0171042-A 24544 27-SEP-2001;
           PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1. .10205
                     /organism="Drosophila sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:7242"
ORIGIN

Alignment Scores:
Pred. No.:                2.19e+03        Length:         10205
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x CQ596786 (1-10205)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       9574 CACAATCACAACCACAAC 9591

RESULT 374
AC013023/c
LOCUS       AC013023               10578 bp    DNA     linear   HTG 03-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC013023
VERSION     AC013023.1  GI:6223068
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 10578)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10212472 by the submitter.
            For further information on this sequence you may e-mail to
            fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .10578
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
ORIGIN

Alignment Scores:
Pred. No.:                2.27e+03        Length:         10578
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       14              Gaps:           0

US-09-858-332G-28 (1-6) x AC013023 (1-10578)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       5715 CACAATCACAATCACAAT 5698

RESULT 375
AC013891
LOCUS       AC013891               11123 bp    DNA     linear   HTG 16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC013891
VERSION     AC013891.1  GI:6437444
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 11123)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10211364 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .11123
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
ORIGIN

Alignment Scores:
Pred. No.:                2.38e+03        Length:         11123
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       14              Gaps:           0

US-09-858-332G-28 (1-6) x AC013891 (1-11123)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        776 CATAATCACAATCATAAT 793

RESULT 376
AF126256/c
LOCUS       AF126256               11283 bp    DNA     linear   BCT 16-MAR-2004
DEFINITION  Aeromonas punctata putative transposase (tnp), RmlB (rmlB), FlmA
            (flmA), FlmB (flmB), NeuA (neuA), FlmD (flmD), NeuB (neuB), Lsg
            (lsg), and Lst (lst) genes, complete cds; and insertion sequence
            IS30 transposase gene, complete cds.
ACCESSION   AF126256
VERSION     AF126256.2  GI:45478495
KEYWORDS    .
SOURCE      Aeromonas punctata (Aeromonas caviae)
  ORGANISM  Aeromonas punctata
            Bacteria; Proteobacteria; Gammaproteobacteria; Aeromonadales;
            Aeromonadaceae; Aeromonas.
REFERENCE   1  (bases 1 to 11283)
  AUTHORS   Gryllos,I., Shaw,J.G., Gavin,R., Merino,S. and Tomas,J.M.
  TITLE     Role of flm locus in mesophilic Aeromonas species adherence
  JOURNAL   Infect. Immun. 69 (1), 65-74 (2001)
   PUBMED   11119490
REFERENCE   2  (bases 1 to 11283)
  AUTHORS   Gryllos,I. and Shaw,J.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-FEB-1999) Molecular and Genetic Medicine, University
            of Sheffield Medical School, Beech Hill Road, Sheffield S10 2RX, UK
REFERENCE   3  (bases 1 to 11283)
  AUTHORS   Tabei,S.M.B., Gryllos,I. and Shaw,J.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAR-2004) Division of Genomic Medicine, School of
            Medicine and Biomedical Sciences, University of Sheffield, Beech
```

```
            Hill Road, Sheffield S10 2RX, UK
  REMARK    Sequence update by submitter
COMMENT     On Mar 16, 2004 this sequence version replaced gi:5616169.
FEATURES             Location/Qualifiers
     source          1. .11283
                     /organism="Aeromonas punctata"
                     /mol_type="genomic DNA"
                     /strain="Sch3"
                     /db_xref="taxon:648"
     gene            273. .1010
                     /gene="tnp"
     CDS             273. .1010
                     /gene="tnp"
                     /note="member of transposase 11 family"
                     /codon_start=1
                     /transl_table=11
                     /product="putative transposase"
                     /protein_id="AAS66621.1"
                     /db_xref="GI:45478497"
                     /translation="MLKGIFNLTLRATQGLLDSLFELMNVPLCAPDYSCVSKRARTVA
                     VAYRQPSKGRITDLVIDSTGLKVFGEGEWKVRKHGAEKRRVWRKLHLAVDPVTHDIVA
                     AEVSLENVHDAEVLPTLLNPLRRKLGRVYADGAYDSKASHQLISRKGATACIPPRKNA
                     GLWEKGHPRNEAVLVMRKEGLAHWKKISGYHRRSLAETAMYRFKQLLAGRISLRNYNG
                     QVGEVMAYVSAINKLNTLGLPVRKPRV"
     gene            1306. .2370
                     /gene="rmlB"
     CDS             1306. .2370
                     /gene="rmlB"
                     /note="putative dTDP-D-glucose-4,6-dehydratase"
                     /codon_start=1
                     /transl_table=11
                     /product="RmlB"
                     /protein_id="AAS66622.1"
                     /db_xref="GI:45478498"
                     /translation="MKILVTGGAGFIGSAVVRHIIRYTRDSVVNLDKLTYAADPASLA
                     SVSTSDRYVFEQVDICNRVEVDRVFTTHQPDAVMHLAAESHVDRSITGPADFIQTNIV
                     GTYMLLEAARAYWNDLDEVRKAAFRFHHVSTDEVYGDLLHPDELTSGAALPFFTETTS
                     YAPSSPYSAAKASSDHLVRAWHRTYGLPVLVTNCSNNYGPWQNAEKLIPQTIAHALAG
                     KPIPLYGRGDQIRDWLYVEDHVRALLLVLTKGRVGETYNIGGHNEKRNLEVVQTICDL
                     LDELVPLSPSPLEGEKPGERVKSYRDQITHVADRPGHDRRYAIDAGKIATELKWLPQE
                     SFESGLYKTVQWYLDSSWCL"
     gene            2521. .3534
                     /gene="flmA"
     CDS             2521. .3534
                     /gene="flmA"
                     /note="putative UDP-GlcNAc C6 dehydratase; C4 reductase"
                     /codon_start=1
                     /transl_table=11
                     /product="FlmA"
                     /protein_id="AAD45656.2"
                     /db_xref="GI:45478496"
                     /translation="MLNNKTVLITGGTGSFGKQFIKTILERYPQVKRIVIFSRDELKQ
                     SELRLNYPQKDYPQLRFFIGDVRDRNRMVQACEGIDVIIHAAAIKQVDTAEYNPTECI
                     RTNVDGAENVIHAALQCGVKEVVALSTDKACAPINLYGATKLTSDKLFTAANNIKGSR
                     NIRFSVVRYGNVMGSRGSVIPFFLKKRAEGVLPITHEEMTRFNISLQDGVNMVMYALE
                     HHLGGEIFVPKIPSYRILDIATAISPECKTKVVGIRPGEKLHEEMITDTDSLNTIDLG
                     RYYAILPSVSFNHTEEDYIKHHSAVKVPFGFKYNSGTNTEWETVESLRELIKTHVDAN
                     FEV"
     gene            3543. .4706
                     /gene="flmB"
     CDS             3543. .4706
                     /gene="flmB"
                     /note="putative pyridoxal-phosphate dependent
                     aminotransferase"
                     /codon_start=1
                     /transl_table=11
                     /product="FlmB"
                     /protein_id="AAD45657.1"
                     /db_xref="GI:5616171"
                     /translation="MIPYGKQEITQQDIDGVVAVLQSDFLTQGPKVPEFEKYLTTHTG
                     AAYALAVNSATSALHIACLALGLGPGDFLWTSPVTFVASANCGLYCGAKVDFVHIDPD
                     TYNLCPKQLEKKLIAANANGTLPKIVVPVHLCGQPCDMEAISRLATEYGFKVIEDASH
                     AIGGSYLNEPIGNCRYSDITVFSFHPVKIITTAEGGAVVTNQKDLADKMALLRSHGIT
                     RDIEQMKGESHGGWYYQQIDLGFNYRMTELQAALGITQIQRLDEFIAARHRLADRYNR
                     LLKRLPIVLPFQLENTHSGLHLYVIRLQLNNISLTHREVFEALRENGIGVNVHYIPVH
                     TQPYYEDMGFKSGDFPEAEQYYREAISLPMFHGMTEEQQDKVVCVLSEILEGK"
     gene            4707. .5393
                     /gene="neuA"
     CDS             4707. .5393
                     /gene="neuA"
                     /note="similar to CMP-NeuNAc synthetase"
                     /codon_start=1
                     /transl_table=11
                     /product="NeuA"
                     /protein_id="AAD45658.1"
                     /db_xref="GI:5616172"
                     /translation="MNIAIIPARGGSKRIPRKNIKPFHSKPMIAWSILAAKKAGCFER
                     IIVSTDDAEIAAVALEYGAEVPFTRPAEIANDYATTGEVISHAINWLINQQGQVPENV
                     CCLYATAPFVEPDDLCQGLELLTFNKECQFVFSATRFSFPIQRAIKLDESGWVSMFHP
                     EYQLTRSQDLEEAYHDAGQFYWGKANAWLNKLPIFAVHTQVVLLPSHRVQDIDTQDDW
                     LRAEKLFTLR"
     gene            5399. .6916
                     /gene="flmD"
     CDS             5399. .6916
                     /gene="flmD"
                     /note="contains glycosyltransferase and acetyltransferase
                     domains"
                     /codon_start=1
                     /transl_table=11
                     /product="FlmD"
                     /protein_id="AAD45659.1"
                     /db_xref="GI:5616173"
                     /translation="MKFFIRTDASRWIGSGHVMRCLMLADGLKSAGHQVTFICRPQDG
                     DLLDLIAARNHNLLALPSLTTVIRPEHGADYSSWLQVPWQQDAADVLESVVYADWLIV
                     DHYGINAQWEREIRLSLNCKVLSIDDLVRPHDADLILDQTLDRTPEAYFECFNISDKQ
                     PVVLSGVAYALLSSQFGKIRNQGILRAPASRQPRILLSFGGIDEPNVTLRVLKALSEF
                     NAVFTVLLSPRAPHYQAVKTWCDDRVNVVHHDFIQDMATLMSEHDLAIGAPGTTTWER
                     ACLGLPNIVIPIAENQQAICQALLAHQAALLVQQSDIESRLIPALHDLLNNYKEYVDN
                     NLELCDGKGVQRVINTIMANTEISGFRLRPVSIDDTLLIYQWQSDPITRRYALNPNPP
                     SWVEHKSWMEKKIQCQQDYFFLIINAESSEPCGAVRLDELGTDNYLISIFVAPNKHGQ
                     GIGLTGLKLLDRCLPKIRIHATVLTENIASQRLFSNAGYQQVAPDKFIREPLWTK"
     gene            6904. .7962
                     /gene="neuB"
     CDS             6904. .7962
                     /gene="neuB"
                     /note="putative pseudaminic acid synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="NeuB"
                     /protein_id="AAD45660.1"
                     /db_xref="GI:5616174"
                     /translation="MDKIIKPFITINGRKIGPDYPPYIIAELSANHNGDINRAFAIME
                     EAKKAGADAIKLQTYTADTITFECDSEEFQIHGGLWDGKNLYQLYKEAQMPWEWHQPL
                     FEKAKELGITIFSSPFDFTAVDLLEDLDAPAYKIASFELIDLPLIKRVAQTGKPMIMS
                     TGMANAEEIAEAIATAQSNGCQELVVLHCVSGYPAPADQYNLRTIADMAERFGVLSGL
                     SDHTIDNATAVAAVALGACLIEKHVTMDRNGGGADDSFSLEPHELAALCKDAKTAWSA
                     LGSVNYTRTEAEKGNVKFRRSLYVIRDIKAGDVLSNDNVRSIRPGFGLAPKYLEQVLG
                     KTAIVDIKRGTPLSFEFI"
     gene            8003. .9265
                     /gene="lsg"
     CDS             8003. .9265
                     /gene="lsg"
                     /note="putative member of Wzx family of O-antigen
                     transporters"
                     /codon_start=1
                     /transl_table=11
                     /product="Lsg"
                     /protein_id="AAS66623.1"
                     /db_xref="GI:45478499"
                     /translation="MSLAKSSLVYLISNIINALIPFLLLPILTRLLSPAEYGQVAMFQ
                     ILVAGLAAFVGLNTVGAANRKFYEHNNSLKALGLYNGACLQILLGSMIIVFILCVLFS
                     ELLVSYLSIPVNWIYSAVLLSSFNFILNLRLGQWQIRGEAIKFGALQISNSLFNMGLS
                     LLLVLSLHHGAQGRIDAQIIAAGLSTVVAVYLLMKDGLVHLYSLRVDYIKQALQFGIP
                     LIPHVFGFFLLSAIDRFVINQKLGLGHAGIYMVAVQLSMAFNIVFDALNKAYVPWLFE
                     ILKRDEMAEKRRVVNYTYLYFIFLLAIAPIPFLIGPWALCFIAGEEFRKAGEIIGWLC
                     LGQIFGGMYLMVTNYIFYAKKNSGLSLITITSGIINVGLLLLLVNSFGLIGAAWAFAI
```

```
                TKFIQFISTWLYSSFVLAMPWLPQGGKL"
     gene       9238. .10185
                /gene="lst"
     CDS        9238. .10185
                /gene="lst"
                /note="putative sugar transferase"
                /codon_start=1
                /transl_table=11
                /product="Lst"
                /protein_id="AAS66624.1"
                /db_xref="GI:45478500"
                /translation="MVTSRWETMNLVVCNTPFQALQIINLVAKGIITNFDFFYFCKNK
                TAQVEYYFEQVKKHAISSQLYISDKRYPYHILDIRRLFKGRYYQSVYSASVDSVFTHS
                ILSFIKFDNFYSFDDGSANLNRNSTYYIEQRGYFKRILFRLTGCKYDLRKTKSIIRSH
                YTVYKNKKNIVSNTIYIDFEFPAQCSDGRKETNDKNSVANVLLGTVYDEVFSDNEVII
                KKLSCFFADKDFYYIPHPRDTHNFFINGMRIDGPEIAESKILSLLKRYETINLYGFGS
                SVQINMSSHRGVINYIFDIGSHEYEYDHDYDYGYEILKI"
     CDS        complement(10240. .11187)
                /codon_start=1
                /transl_table=11
                /product="transposase"
                /protein_id="AAS66625.1"
                /db_xref="GI:45478501"
                /translation="MKYQQLTEGLRYQIACLRDHGLSQARIAKQLGVHPSTISRELRR
                NCAGQAYQPQQAHQLATARRLTSQKRTVPSDTIDFVCMALVADWSPEQISGVGDMINC
                PVSHEWIYRYVARDKANGGTLYRHLRQGHKRYRKGKNSKRSVIPNPVSIDERPAIVNE
                RSRVGDWEVDTVLGKQGSGAIVSLLERKSRLYLVQYVPSKTASAVADTIISMLKPYQA
```

Alignment Scores:

Pred. No.:	2.41e+03	Length:	11283
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0

US-09-858-332G-28 (1-6) x AF126256 (1-11283)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      10168 CATAACCATAATCATAAT 10151
```

```
RESULT 377
AB209488/c
LOCUS       AB209488               11401 bp    mRNA    linear   PRI 31-MAR-2005
DEFINITION  Homo sapiens mRNA for large conductance calcium-activated potassium
            channel subfamily M alpha member 1 variant protein.
ACCESSION   AB209488
VERSION     AB209488.1  GI:62088555
KEYWORDS    FLI_CDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Totoki,Y., Toyoda,A., Takeda,T., Sakaki,Y., Tanaka,A., Yokoyama,S.,
            Ohara,O., Nagase,T. and F.Kikuno,R.
  TITLE     None Title
  JOURNAL   Published Only in Database (2005)
REFERENCE   2  (bases 1 to 11401)
  AUTHORS   Totoki,Y., Toyoda,A., Takeda,T., Sakaki,Y., Tanaka,A., Yokoyama,S.,
            Ohara,O., Nagase,T. and F.Kikuno,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-2005) Osamu Ohara, Kazusa DNA Research Institute,
            Department of Human Gene Research; 2-6-7 Kazusa-kamatari, Kisarazu,
            Chiba, 292-0818, Japan (E-mail:cdnainfo@kazusa.or.jp,
            URL:http://protein.gsc.riken.go.jp/, Tel:81-438-52-3930,
            Fax:81-438-52-3931)
COMMENT     This work was supported in part by the National Project on Protein
            Structural and Functional Analysis , Ministry of Education, Culture
            , Sports, Science and Technology of Japan. Totoki Y, Toyoda A,
            Takeda T, Sakaki Y, Tanaka A, Yokoyama S. RIKEN Genomic Sciences
            Center, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan.
            e-mail: aktanaka@postman.riken.go.jp
            URL: http://protein.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1. .11401
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="ff02525"
                     /tissue_type="brain"
                     /note="this clone is also named as hsk003001531
                     vector:pBluescriptII SK plus"
     gene            <1. .11401
                     /gene="large conductance calcium-activated potassium
                     channel subfamily M alpha member 1 variant"
     CDS             <1. .1773
                     /gene="large conductance calcium-activated potassium
                     channel subfamily M alpha member 1 variant"
                     /note="Start codon is not identified."
                     /codon_start=1
                     /evidence=not_experimental
                     /product="large conductance calcium-activated potassium
                     channel subfamily M alpha member 1 variant"
                     /protein_id="BAD92725.1"
                     /db_xref="GI:62088556"
                     /translation="GTLKPVDEKEEAVAAEVGWMTSVKDWAGVMISAQTLTGRVLVVL
                     VFALSIGALVIYFIDSSNPIESCQNFYKDFTLQIDMAFNVFFLLYFGLRFIAANDKLW
                     FWLEVNSVVDFFTVPPVFVSVYLNRSWLGLRFLRALRLIQFSEILQFLNILKTSNSIK
                     LVNLLSIFISTWLTAAGFIHLVENSGDPWENFQNNQALTYWECVYLLMVTMSTVGYGD
                     VYAKTTLGRLFMVFFILGGLAMFASYVPEIIELIGNRKKYGGSYSAVSGRKHIVVCGH
                     ITLESVSNFLKDFLHKDRDDVNVEIVFLHNISPNLELEALFKRHFTQVEFYQGSVLNP
                     HDLARVKIESADACLILANKYCADPDAEDASNIMRVISIKNYHPKIRIITQMLQYHNK
                     AHLLNIPSWNWKEGDDAICLAELKLGFIAQSCLAQGLSTMLANLFSMRSFIKIEEDTW
                     QKYYLEGVSNEMYTEYLSSAFVGLSFPTVCELCFVKLKLLMIAIEYKSANRESRSRKR
                     ILINPGNHLKIQEGTLGFFIASDAKEVKRAFFYCKACHDDITDPKRIKKCGCKRLEDE
                     QPSTLSPKKKATEWRHAELTQHLA"
ORIGIN
```

Alignment Scores:

Pred. No.:	2.44e+03	Length:	11401
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	8	Gaps:	0

US-09-858-332G-28 (1-6) x AB209488 (1-11401)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      11206 CATAATCATAATCATAAT 11189
```

```
RESULT 378
CS133215
LOCUS       CS133215               11587 bp    DNA     linear   PAT 02-AUG-2005
DEFINITION  Sequence 37 from Patent WO2005061736.
ACCESSION   CS133215
VERSION     CS133215.1  GI:71792282
KEYWORDS    .
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1
  AUTHORS   Tononi,G.
  TITLE     Sleep genes in drosophila and their use for the screening,
            diagnosis and therapy of sleep disorders
  JOURNAL   Patent: WO 2005061736-A 37 07-JUL-2005;
            Wisconsin Alumni Research Foundation (US)
FEATURES             Location/Qualifiers
     source          1. .11587
                     /organism="Drosophila melanogaster"
```

```
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:7227"
ORIGIN

Alignment Scores:
Pred. No.:                2.48e+03        Length:          11587
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       6               Gaps:            0

US-09-858-332G-28 (1-6) x CS133215 (1-11587)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        734 CATAATCACAATCACAAT 751

RESULT 379
CQ586445
LOCUS       CQ586445               13587 bp    DNA     linear   PAT 02-FEB-2004
DEFINITION  Sequence 14203 from Patent WO0171042.
ACCESSION   CQ586445
VERSION     CQ586445.1  GI:41646276
KEYWORDS    .
SOURCE      Drosophila sp.
  ORGANISM  Drosophila sp.
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1
  AUTHORS   Venter,J.C., Adams,M., Li,P.W. and Myers,E.W.
  TITLE     Detection kits, such as nucleic acid arrays, for detecting the
            expression of 10,000 or more Drosophila genes and uses thereof
  JOURNAL   Patent: WO 0171042-A 14203 27-SEP-2001;
            PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1. .13587
                     /organism="Drosophila sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:7242"
ORIGIN

Alignment Scores:
Pred. No.:                2.89e+03        Length:          13587
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       6               Gaps:            0

US-09-858-332G-28 (1-6) x CQ586445 (1-13587)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1734 CATAATCACAATCACAAT 1751

RESULT 380
AC115588/c
LOCUS       AC115588               14287 bp    DNA     linear   INV 12-MAR-2003
DEFINITION  Dictyostelium discoideum chromosome 2 map
            complement(5515123-5500836) strain AX4, complete sequence.
ACCESSION   AC115588
VERSION     AC115588.2  GI:28828567
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 14287)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 14287)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 14287)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 14287)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On Mar 4, 2003 this sequence version replaced gi:19569958.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES             Location/Qualifiers
     source          1. .14287
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /chromosome="2"
                     /map="complement(5515123-5500836)"
     CDS             join(3405. .3506,3651. .4061)
                     /note="GeneID exon scores (in order of location ranges):
                     3.01, 46.26 - GSCJ_ID dd_00471"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO51172.1"
                     /db_xref="GI:28828568"
                     /translation="MKYLLCLDGSKKARQAVESTLKFVNKGDELYFVSVYPPHPNPNE
                     NNEPIDSSTIAKMVIEMFSGVPDLVDPDEEKLAKNEQIINSNHKKVNALEEFHYFNNK
                     NLSNDKCKIHYLLVASNDISGSITDVANKYGINCVVLGSTGTSLEKLLIGSVSSKVLR
                     TSHCSVLISP"
     CDS             complement(join(4301. .6118,6213. .6410,6500. .7774))
                     /note="GeneID exon scores (in order of location ranges):
                     188.52, 16.07, 130.75 - GSCJ_ID dd_00472"
                     /codon_start=1
                     /product="similar to Dictyostelium discoideum (Slime
                     mold). GrfA (Fragment)"
                     /protein_id="AAO51173.1"
                     /db_xref="GI:28828569"
                     /translation="MENNNSNNNNNNLSTLRKSNINGGNSNIVQSCLSFLHRSLSIK
                     KWIEKILGEKLSEDHHDDNDGLLLFGKLMNGVVLCKLMSCLKSNSIPKIHTGESILNF
                     KIRENLCFYIQALEDIGISRHFQFQVPDLLEKKNLLKVLESLEVLADDCTDFEKPIEL
                     SDEESLLIQFTDKQLSDTEQILKKLNVTHIKRQSLKKLTSNEPKRFTTSGTTLSHSTG
                     SRLTFGGGSSNSSTGNSNGSSSGGGSSLSHSGGAPHSSSTPSSLSSSSNGGSGLTQSG
                     GGSRTFQTRIVSTPANVKPKAPLSVAQILDKARPFEKRWIKIQAFFRGHLARKAYQKR
                     VRFAAYRHNIVKELLATEQVYIDSLDHVVAHYFNPMREESSKLNIKLDHFGLLISNLE
                     VIINYNKNLLQNIKPKVENWSHSQTIGDIFEQFTLYLKVYTQYVKEYSIFFEAINNLR
                     KSNTKFESFIVEKEYNDNYKSIGTYLILPIQRIPRYTLLLTELLKNTWSDHLDYNNLQ
                     LSLKKMQEVALYVNEKKREAENIAKVTEIQNNFIGKFENLAEPHRRFVFEGHLSVIGP
                     GGKESQRVVYLFNDVLISTKPITSGLVKKKVSLKVKESIRMNLVTIKSKSNNNNNNNI
                     ENSENNNNNNSNNNIANNNTPTKISSTPKSSSASSLLSNSTPSSPSPSTTNTTTNTPQ
                     PTPTPISLSSSMPSQPTKPTTLKPVPPPRSQATFNVTSPTPNPTTTNPTTTNPTTTNP
                     TTTTTTTTNKPTIGVKTSSSSITISSTNKEVVESTKPRSSTTVVNNSTNINIHNQIRQ
                     PTPIASVTASSSQERLIIEEKMKKQNQTIELIDTFGTCVLKLLCSSVKEKDQWVSEIK
                     KVYEDLDNKKIINDEAMKRSQERAGLAKAALSQQYATLRVKGRLCDLNTLKTSINQEE
                     DGSNNSNSNNSSINNSPTTSSPIVNSVNGSGSGSGGGGGSEQQDGSTSNTTNTSGERE
                     FNVLRRQTLSARRSQNNATPPPPQSQSTTTANTNTTTTNDSASDHSDTDEKKNEKSHS
```

```
                     KTGTGWFSSLRKKKKHPSIQDTFGLDTSPITPSDIKSSTSTTTPPTSTSTSTPPQSNT
                     VILKQDNL"
     CDS             join(8952. .9008,9185. .9688)
                     /note="GeneID exon scores (in order of location ranges):
                     4.79, 52.78 - GSCJ_ID dd_00473"
                     /codon_start=1
                     /product="similar to transport protein particle component
                     Bet3p-like protein; protein id: At5g54750.1, supported by
                     cDNA: 122866. [Arabidopsis thaliana]"
                     /protein_id="AAO51174.1"
                     /db_xref="GI:28828570"
                     /translation="MSKKYDKLGNDVFNKVEKINSELFTLTYGALVTQLIKDYEDIEQ
                     VNIKLEQMGYNIGIRLIEEFLAKSGIGRCSDFIETAEVIAKVGFKMFLGVNAHVGDWD
                     ADRKEFHLTIEDNPLIDFVELPDQYKHKLYYSNILCGVMRGALEMVQMKVKCTFVKCT
                     LSDDSTSEIKVVLEEVLSDMIPVGYD"
     CDS             join(10157. .10438,10618. .11769)
                     /note="GeneID exon scores (in order of location ranges):
                     13.35, 117.19 - GSCJ_ID dd_00474"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO51175.1"
                     /db_xref="GI:28828571"
                     /translation="MNNDKEINSEKSKRAINHFSNKKEIKNSDRRGILFSKLPSQQLN
                     PDRYDNLMTFWNNSLIDISKSCNILIFTPKLLSTYFTVENVSPIYLPLILNEMIKTKF
                     IVKYEEYIKDYGWSKWVWTKMVVQPFQYYTGLSTPNTETEKNVKFILPEMIKDKAEQL
                     YQHQLKNMNSTTDNIISYINLEKSIKDWYITREELDLLLLVLFRDSKVLILTNNKIKN
                     INNNNNGEDRIGIKFAFDGEKVQPIQETDFGILKLQTTYETLKQQESKLLTDIEEISN
                     TIKESIRIKQKNHALLQLKKKKLLESILEKRATASTNIHEILFSIESAKSNQQIIESL
                     CTGVSTLKKVNQEISVDQVDSILDDYQDAITNQREIDDAMKSGFNSVESLSSADIDED
                     QLEKELDQMLQDHLTLEKEEKQKQKQIEKEKQQQQQPPTSQFNPNLPIPLKNEEDELL
                     KELESLSVTSNPLPKQDENKQKTSELI"
     CDS             complement(join(11853. .12979,13123. .>13456))
                     /note="GeneID exon scores (in order of location ranges):
                     70.91, 22.33 - GSCJ_ID dd_00475"
                     /codon_start=1
                     /product="similar to Mus musculus (Mouse). Adult male
                     hypothalamus cDNA, RIKEN full-length enriched library,
                     clone:A230106L22 product:hyp"
                     /protein_id="AAO51176.1"
                     /db_xref="GI:28828572"
                     /translation="VEFTLLEDDVYFHDYIADFIPLFVLSILLEFVINVFFKKKKNYH
                     SLSDSLSSMGTGISTVVLERIVPYTILSSIEFVGAVYIYKHYRFFDVSESSILAWIIC
                     LLAVDFVYYWFHRFAHEINFMWATHVTHHSSDKYNLTTALRQSIFQMYFSWVLFLPIG
                     FFIPPGIYVFHKQFNTINQFWIHTQLIGKLAWPIEFIMNTPSHHRVHHGRNPKYLDKN
                     YAGTLIIWDRLFGTFQEEDTITNTTTTTTNTNGENNNNNNNNNNNNNNNNNNNENQR
                     VFYGLVHQFESSDPTWSQVHHWFEIYEKQKQYNSVKEKLKSIFYGPGWKKGDTNRMGD
                     LSSIPIPTEKDDQLRIKASLSTSLNIYLFIHYFLASFLALSLRSLFEDKIGSANVTCS
                     FIFVYFGLTSFGAIFDRKWYAMILEHFRLWSFILYTLYLESTNPIISQQFQIISLLRY
                     IYFVSSLFLILKKLSPLSILSNYLKSSKKLLNNKLD"
ORIGIN

Alignment Scores:
Pred. No.:                3.03e+03        Length:          14287
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       2               Gaps:            0

US-09-858-332G-28 (1-6) x AC115588 (1-14287)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1180 CACAACCACAACCACAAC 1163

RESULT 381
CS018257
LOCUS       CS018257               14720 bp    DNA     linear   PAT 23-FEB-2005
DEFINITION  Sequence 24 from Patent WO2005012534.
ACCESSION   CS018257
VERSION     CS018257.1  GI:60220038
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Alphey,L.
  TITLE     Expression systems for insect pest control
  JOURNAL   Patent: WO 2005012534-A 24 10-FEB-2005;
            Oxitec Limited (GB)
FEATURES             Location/Qualifiers
     source          1. .14720
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="pLA1038"
ORIGIN

Alignment Scores:
Pred. No.:                3.12e+03        Length:          14720
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       6               Gaps:            0

US-09-858-332G-28 (1-6) x CS018257 (1-14720)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       6231 CACAACCATAATCACAAT 6248

RESULT 382
AC013163
LOCUS       AC013163               14923 bp    DNA     linear   HTG 03-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC013163
VERSION     AC013163.1  GI:6223168
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 14923)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10213807 by the submitter.
            For further information on this sequence you may e-mail to
            fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .14923
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
ORIGIN

Alignment Scores:
Pred. No.:                3.16e+03        Length:          14923
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       14              Gaps:            0

US-09-858-332G-28 (1-6) x AC013163 (1-14923)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      12875 CACAATCACAATCACAAT 12892
```

```
RESULT 383
AY458594
LOCUS       AY458594               15983 bp    DNA     circular VRT 06-JUL-2005
DEFINITION  Ichthyophis bannanicus mitochondrion, complete genome.
ACCESSION   AY458594
VERSION     AY458594.1  GI:46981621
KEYWORDS    .
SOURCE      mitochondrion Ichthyophis bannanicus
  ORGANISM  Ichthyophis bannanicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Gymnophiona; Ichthyophiidae; Ichthyophis.
REFERENCE   1  (bases 1 to 15983)
  AUTHORS   Zhang,P., Zhou,H., Chen,Y.Q., Liu,Y.F. and Qu,L.H.
  TITLE     Mitogenomic Perspectives on the Origin and Phylogeny of Living
            Amphibians
  JOURNAL   Syst. Biol. 54 (3), 391-400 (2005)
REFERENCE   2  (bases 1 to 15983)
  AUTHORS   Zhang,P., Chen,Y.Q., Zhou,H. and Qu,L.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-NOV-2003) Biotechnology Research Center, Life Science
            School, Zhongshan University, GuangZhou, Guangdong 510275, China
FEATURES             Location/Qualifiers
     source          1. .15983
                     /organism="Ichthyophis bannanicus"
                     /organelle="mitochondrion"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:8453"
     tRNA            1. .69
                     /product="tRNA-Phe"
     rRNA            70. .1009
                     /product="12S ribosomal RNA"
     tRNA            1010. .1076
                     /product="tRNA-Val"
     rRNA            1077. .2634
                     /product="16S ribosomal RNA"
     tRNA            2636. .2709
                     /product="tRNA-Leu"
                     /note="codons recognized: UUR"
     gene            2710. .3672
                     /gene="ND1"
     CDS             2710. .3672
                     /gene="ND1"
                     /codon_start=1
                     /transl_table=2
                     /product="NADH dehydrogenase subunit 1"
                     /protein_id="AAT07873.1"
                     /db_xref="GI:46981622"
                     /translation="MLSMILSPLTLIIPILLSVAFLTLIERKILGYMQLRKGPNIVGP
                     TGLLQPIADGLKLFIKEPVRPSTSSQMMFIMTPMLALTLAMTLWSPMPMPSPILNMNL
                     GLLFILSLSSLAVYSILGSGWASNSKYALIGALRAVAQTISYEVTLGLIMLSTILLSG
                     SFTLYNFMYAQEPTWLLLPTWPMAAMWFISTLAETNRAPFDLTEGESELVSGFNVEYA
                     GGPFALFFLAEYSNILMMNTLSTIIFFSPTQNFLSPETTTTNLMVKATLLSMIFLWVR
                     ASYPRFRYDQLMHLVWKNFLPLVLATTLLYISMPISFSGVPPQL"
     tRNA            3673. .3741
                     /product="tRNA-Ile"
     tRNA            complement(3741. .3811)
                     /product="tRNA-Gln"
     tRNA            3811. .3878
                     /product="tRNA-Met"
     gene            3879. .4916
                     /gene="ND2"
     CDS             3879. .4916
                     /gene="ND2"
                     /codon_start=1
                     /transl_table=2
                     /product="NADH dehydrogenase subunit 2"
                     /protein_id="AAT07874.1"
                     /db_xref="GI:46981623"
                     /translation="MNPYALSIMLSSIATGTIIVVSSSHWLLVWIGLEINTLAIIPLM
                     SQQLHPRSTEATMKYFLTQAIASALILFSATTNAWLTNQWQMTELLPLQTNMITIALA
                     MKLGLAPMHFWLPEVLQGIDLTTGLILSTWQKIAPMMVIYMTYNTLNPHLLTPMAILS
                     TIIGGWGGLNQTQLRKMMAYSSIAHLGWMMAVATMCPNLMLLNFMVYLIMTSSMFLTM
                     KYLKTTSINLLSVSWSSTPTMAAISSLTLLSLGGLPPTSGFMPKWLILQELTKQHTTI
                     TATMLAMSSLLSLFFYLRLTYSMTLTNSPNTSNATTTWRQKSNQASMPYSLVTIMTIL
                     LLPITPNLLPQ"
     tRNA            4915. .4983
                     /product="tRNA-Trp"
     tRNA            complement(4985. .5053)
                     /product="tRNA-Ala"
     tRNA            complement(5054. .5126)
                     /product="tRNA-Asn"
     rep_origin      complement(5127. .5162)
                     /note="L-strand origin of replication"
     tRNA            complement(5163. .5223)
                     /product="tRNA-Cys"
     tRNA            complement(5225. .5294)
                     /product="tRNA-Tyr"
     gene            5296. .6846
                     /gene="COI"
     CDS             5296. .6846
                     /gene="COI"
                     /codon_start=1
                     /transl_table=2
                     /product="cytochrome c oxidase subunit I"
                     /protein_id="AAT07875.1"
                     /db_xref="GI:46981624"
                     /translation="MMMTRWLFSTNHKDIGTLYLIFGAWAGMVGTALSLLIRAELSQP
                     GTLLGNDQIYNVIVTAHAFVMIFFMVMPIMIGGFGNWLVPLMIGAPDMAFPRMNNMSF
                     WLLPPSFLLLLASSGVEAGAGTGWTVYPPLAGNLAHAGASVDLTIFSLHLAGVSSILG
                     AINFITTIVNMKPPAVSQYQTPLFVWSVLITAVLLLLSLPVLAAGITMLLTDRNLNTT
                     FFDPAGGGDPILYQHLFWFFGHPEVYILILPGFGMISHIVTYYSGKKEPFGYMGMVWA
                     MMSIGLLGFIVWAHHMFTVGMNVDTRAYFTSATMIIAIPTGVKVFSWLATLHGGAIKW
                     DAAMLWSLGFIFLFTVGGLTGIVLANSSLDIVLHDTYYVVAHFHYVLSMGAVFAIMGG
                     FIHWFPLFSGYTLHNTWTKIHFGVMFAGVNMTFFPQHFLGLAGMPRRYSDYPDAYTLW
                     NSVSSIGSLISLVAVIMMMFIIWEAFSAKREVKTTELTMTNVEWLHGCPPPYHTYEEP
                     AFVQVQKG"
     tRNA            complement(6835. .6907)
                     /product="tRNA-Ser"
                     /note="codons recognized: UCN"
     tRNA            6911. .6979
                     /product="tRNA-Asp"
     gene            6980. .7672
                     /gene="COII"
     CDS             6980. .7672
                     /gene="COII"
                     /codon_start=1
                     /transl_table=2
                     /product="cytochrome c oxidase subunit II"
                     /protein_id="AAT07876.1"
                     /db_xref="GI:46981625"
                     /translation="MAQPSQLGFQDAASPVMEELLNFHDHTLMIVFLISTLVLYTITI
                     MMTTKLTNTNTMDAQEVEMIWTVLPAIIMIAIALPSLRILYLMDEINNPHVTVKAIGH
                     QWYWTYEFTDFNMLSFDSYMTPSNHLSPGEFRLLEVDNRMVVPMELPVRMMISAEDVL
                     HSWAIPSLGVKTDAIPGRLNQTTLTTSRPGVYYGQCSEICGANHSFMPIVLEATPYFE
                     FLTWTTMVETSL"
     tRNA            7665. .7735
                     /product="tRNA-Lys"
     gene            7737. .7901
                     /gene="ATP8"
     CDS             7737. .7901
                     /gene="ATP8"
                     /codon_start=1
                     /transl_table=2
                     /product="ATP synthase subunit 8"
                     /protein_id="AAT07877.1"
                     /db_xref="GI:46981626"
                     /translation="MPQLNPSPWFLIFLLSWVVLTTSIAVKTMKHQPVNDISTENQTK
                     AQSPWNWPWH"
     gene            7892. .8575
                     /gene="ATP6"
     CDS             7892. .8575
                     /gene="ATP6"
                     /codon_start=1
                     /transl_table=2
                     /product="ATP synthase subunit 6"
```

```
                     /protein_id="AAT07878.1"
                     /db_xref="GI:46981627"
                     /translation="MTLALFDQFMISTFMCIPLILLPMTLPWILITSQHKKWTVNRLT
                     AIQSWFLFYSTKQLMATMNMKGQKWALLLISLMTFLILTNLLGLLPYTFTPTTQLSMN
                     MALAIPMWLATVLLGMRHQMTTSLGHLLPEGTPTLLIPPLVLIETISLFIRPLALGVR
                     LTANLTAGHLLIQLISTASMLLMPTMPTISLATTTILLLLTVLELAVAMIQAYVFVLL
                     LTLYLQENI"
     gene            8575. .9358
                     /gene="COIII"
     CDS             8575. .9358
                     /gene="COIII"
                     /note="TAA stop codon is completed by the addition of 3' A
                     residues to the mRNA"
                     /codon_start=1
                     /transl_except=(pos:9358,aa:TERM)
                     /transl_table=2
                     /product="cytochrome c oxidase subunit III"
                     /protein_id="AAT07879.1"
                     /db_xref="GI:46981628"
                     /translation="MAHQSHAFHMVDPSPWPLTGAIAALMLTSGLATWFHFNSTKILL
                     ISLVVLLLTMYQWWRDIIRESTFQGHHTQPVQKGLRYGMILFITSEVFFFLGFFWAFY
                     RASLAPTIELGECWPPTGILPLDPFEVPLLNTAVLLASGVTVTWAHHSIMLGNRKEAI
                     QSLMLTILLGLDFTLLQAMEYYEASFTISDGIYGSTFFVATGFHGLHVIIGSTFLTIC
                     LMRQIMFHFTSQHHFGFEAAAWYWHFVDVVWLFLYVSIYWWGS"
     tRNA            9359. .9428
                     /product="tRNA-Gly"
     gene            9430. .9774
                     /gene="ND3"
     CDS             9430. .9774
                     /gene="ND3"
                     /codon_start=1
                     /transl_table=2
                     /product="NADH dehydrogenase subunit 3"
                     /protein_id="AAT07880.1"
                     /db_xref="GI:46981629"
                     /translation="MTMVMMLTTTALITVLMFIAFWLPNTNQDMEKLSPYECGFDPLG
                     SARLPFSMRFFLIAILFLLFDLEIALLLPTPWANQLNTPSTSIFLAMLILLLTLGLI
                     YEWLQGGLEWAE"
     tRNA            9773. .9840
                     /product="tRNA-Arg"
     gene            9841. .10137
                     /gene="ND4L"
     CDS             9841. .10137
                     /gene="ND4L"
                     /codon_start=1
                     /transl_table=2
                     /product="NADH dehydrogenase subunit 4L"
                     /protein_id="AAT07881.1"
                     /db_xref="GI:46981630"
                     /translation="MTLTTYTLYSSFMLGMTGMILHRTHLLSALICLEGMMLSLFITL
                     SLWCVQAESTSYTLTPMILLTFSACEAGTGLAILVATIRTHGTDTMQNLNLLQC"
     gene            10131. .11495
                     /gene="ND4"
     CDS             10131. .11495
                     /gene="ND4"
                     /codon_start=1
                     /transl_table=2

Alignment Scores:
Pred. No.:              3.37e+03        Length:         15983
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     5               Gaps:           0

US-09-858-332G-28 (1-6) x AY458594 (1-15983)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       7099 CACAATCACAATCATAAT 7116

RESULT 384
AF362076S1/c
LOCUS       AF362076S1              16987 bp    DNA     linear   ROD 16-MAY-2001
DEFINITION  Mus musculus alpha 1C/A-adrenergic receptor gene, exon 1.
ACCESSION   AF362076
VERSION     AF362076.1  GI:14091080
KEYWORDS    .
SEGMENT     1 of 3
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 16987)
  AUTHORS   O'Connell,T.D., Rokosh,D.G. and Simpson,P.C.
  TITLE     Cloning and characterization of the mouse alpha1C/A-adrenergic
            receptor gene and analysis of an alpha1C promoter in cardiac
            myocytes: role of an MCAT element that binds transcriptional
            enhancer factor-1 (TEF-1)
  JOURNAL   Mol. Pharmacol. 59 (5), 1225-1234 (2001)
   PUBMED   11306707
REFERENCE   2  (bases 1 to 16987)
  AUTHORS   O'Connell,T.D., Rokosh,D.G. and Simpson,P.C. Jr.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAR-2001) Cardiology, VAMC, 4150 Clement St., San
            Francisco, CA 94121, USA
FEATURES             Location/Qualifiers
     source          1. .16987
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="129/Sv"
                     /db_xref="taxon:10090"
     exon            4478. .5957
                     /number=1
     misc_feature    5066
                     /note="translation initiation site"
ORIGIN

Alignment Scores:
Pred. No.:              3.58e+03        Length:         16987
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     9               Gaps:           0

US-09-858-332G-28 (1-6) x AF362076S1 (1-16987)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       2208 CACAACCACAACCACAAC 2191

RESULT 385
AL672209/c
LOCUS       AL672209                17284 bp    DNA     linear   ROD 09-FEB-2005
DEFINITION  Mouse DNA sequence from clone RP23-296N19 on chromosome 11,
            complete sequence.
ACCESSION   AL672209
VERSION     AL672209.7  GI:20068771
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 17284)
  AUTHORS   Pearce,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-FEB-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 8, 2002 this sequence version replaced gi:19309645.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
```

Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep ----------
Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vega@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.
Sequence from the Mouse Genome Sequencing Consortium whole genome shotgun may have been used to confirm this sequence. Sequence data from the whole genome shotgun alone has only been used where it has a phred quality of at least 30.
RP23-296N19 is from the RPCI-23 Mouse BAC Library constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6.

FEATURES Location/Qualifiers
source 1. .17284
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="11"
/clone="RP23-296N19"
/clone_lib="RPCI-23"

ORIGIN

Alignment Scores:

Pred. No.:	3.64e+03	Length:	17284
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

US-09-858-332G-28 (1-6) x AL672209 (1-17284)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1278 CACAACCACAACCACAAC 1261
```

RESULT 386
U23451
LOCUS U23451 17986 bp DNA linear INV 22-SEP-2004
DEFINITION Caenorhabditis elegans cosmid ZK75, complete sequence.
ACCESSION U23451
VERSION U23451.1 GI:733560
KEYWORDS HTG.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 17986)
AUTHORS .
CONSRTM WormBase Consortium
TITLE Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium
JOURNAL Science 282 (5396), 2012-2018 (1998)
PUBMED 9851916
REFERENCE 2 (bases 1 to 17986)
AUTHORS Kirsten,J.
TITLE The sequence of C. elegans cosmid ZK75
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 17986)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (13-JUL-1995) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
REFERENCE 4 (bases 1 to 17986)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (29-MAY-2002) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
REFERENCE 5 (bases 1 to 17986)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (22-NOV-2002) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
REFERENCE 6 (bases 1 to 17986)
AUTHORS .
CONSRTM WormBase Consortium
TITLE Direct Submission
JOURNAL Submitted (22-SEP-2004) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
COMMENT Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University
St. Louis , MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 IRQ, England
email: submissions@watson.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

For a graphical representation of this clone sequence and its analysis see:
http://www.wormbase.org/db/seq/sequence?name=ZK75;class=Sequence

NEIGHBORING CLONE INFORMATION

The 5' clone is B0034, 200 bp overlap; the 3' clone is ZK84, 400 bp overlap. Actual start of this clone is at base position 1 of ZK75; actual end is at 22444 of ZK84.

NOTES:

Coding seqences below are the result of integration and manual review of the following data : computer analysis using the program Genefinder (P. Green and L. Hillier, personal communication), the large scale EST projects of Yuji Kohara (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C. elegans ORFeome cloning project (http://worfdb.dfci.harvard.edu/), similarity to other proteins from BlastX analyses (http://blast.wustl.edu/), sequence conservation with C. briggsae using Jim Kent's WABA alignment program (Genome Research 10:1115-1125, 2000), individual C. elegans GenBank submissions, and personal communications with C. elegans researchers. tRNAs are predicted using the program tRNAscan-SE (Lowe, T.M. and Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).
FEATURES Location/Qualifiers
source 1. .17986
/organism="Caenorhabditis elegans"
/mol_type="genomic DNA"
/strain="Bristol N2"
/db_xref="taxon:6239"

```
                     /chromosome="II"
                     /clone="ZK75"
     gene            9642. .10286
                     /gene="ins-2"
                     /locus_tag="ZK75.2"
     CDS             join(9642. .9867,10192. .10286)
                     /gene="ins-2"
                     /locus_tag="ZK75.2"
                     /standard_name="ZK75.2"
                     /note="contains similarity to Pfam domain PF03488
                     (Nematode insulin-related peptide beta type); coded for by
                     the following C. elegans cDNAs: OSTR095A9_1"
                     /codon_start=1
                     /product="Insulin related protein 2"
                     /protein_id="AAC46745.1"
                     /db_xref="GI:733561"
                     /db_xref="WormBase:ZK75.2"
                     /translation="MNAIIFCLLFTTVTATYEVFGKGIEHRNEHLIINQLDIIPVEST
                     PTPNRASRVQKRLCGRRLILFMLATCGECDTDSSEDLSHICCIKQCDVQDIIRVCCPN
                     SFRK"
     gene            complement(11412. .11816)
                     /gene="ins-3"
                     /locus_tag="ZK75.3"
     CDS             complement(join(11412. .11497,11579. .11816))
                     /gene="ins-3"
                     /locus_tag="ZK75.3"
                     /standard_name="ZK75.3"
                     /note="contains similarity to Pfam domain PF03488
                     (Nematode insulin-related peptide beta type); coded for by
                     the following C. elegans cDNAs: OSTF095A10_1,
                     OSTF095A10_2, OSTR095A10_1"
                     /codon_start=1
                     /product="Insulin related protein 3"
                     /protein_id="AAC46746.1"
                     /db_xref="GI:733562"
                     /db_xref="WormBase:ZK75.3"
                     /translation="MKLSVVLALFIIFQLGAASLMRNWMFDFEKELEHDYDDSEIGFH
                     NIHSLMARSRRGDKVKICGTKVLKMVMVMCGGECSSTNENIATECCEKMCTMEDITTK
                     CCPSR"
     gene            17500. .17878
                     /gene="ins-4"
                     /locus_tag="ZK75.1"
     CDS             join(17500. .17734,17793. .17878)
                     /gene="ins-4"
                     /locus_tag="ZK75.1"
                     /standard_name="ZK75.1"
                     /note="contains similarity to Pfam domain PF03488
                     (Nematode insulin-related peptide beta type); coded for by
                     the following C. elegans cDNAs: OSTF095B8_1, OSTF113D9_1,
                     OSTR095B8_1, OSTR113D9_1"
                     /codon_start=1
                     /product="Insulin related protein 4"
                     /protein_id="AAC46744.1"
                     /db_xref="GI:733563"
                     /db_xref="WormBase:ZK75.1"
                     /translation="MFSFFTYFLLSALLLSASCRQPSMDTSKADRILREIEMETELEN
                     QLSRARRVPAGEVRACGRRLLLFVWSTCGEPCTPQEDMDIATVCCTTQCTPSYIKQAC
                     CPEK"
ORIGIN

Alignment Scores:
Pred. No.:              3.78e+03        Length:         17986
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-858-332G-28 (1-6) x U23451 (1-17986)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       2743 CATAATCATAATCATAAT 2760
```

```
RESULT 387
U23451/c
LOCUS       U23451                 17986 bp    DNA     linear   INV 22-SEP-2004
DEFINITION  Caenorhabditis elegans cosmid ZK75, complete sequence.
ACCESSION   U23451
VERSION     U23451.1  GI:733560
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 17986)
  AUTHORS   .
  CONSRTM   WormBase Consortium
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
  JOURNAL   Science 282 (5396), 2012-2018 (1998)
   PUBMED   9851916
REFERENCE   2  (bases 1 to 17986)
  AUTHORS   Kirsten,J.
  TITLE     The sequence of C. elegans cosmid ZK75
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 17986)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-1995) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   4  (bases 1 to 17986)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAY-2002) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   5  (bases 1 to 17986)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2002) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   6  (bases 1 to 17986)
  AUTHORS   .
  CONSRTM   WormBase Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-2004) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
COMMENT     Submitted by:
                  Genome Sequencing Center
                  Department of Genetics, Washington University
                  St. Louis , MO  63110, USA, and
                  Sanger Centre, Hinxton Hall
                  Cambridge CB10 IRQ, England
                  email: submissions@watson.wustl.edu and jes@sanger.ac.uk

            NOTICE:  This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate chemistry
            or covered by high quality data (i.e., phred quality >= 30); an
            attempt was made to resolve all sequencing problems, such as
            compressions and repeats; all regions were covered by sequence from
            more than one m13 subclone.

            For a graphical representation of this clone sequence and its
            analysis see:
            http://www.wormbase.org/db/seq/sequence?name=ZK75;class=Sequence

                        NEIGHBORING CLONE INFORMATION
```

The 5' clone is B0034, 200 bp overlap; the 3' clone is ZK84, 400 bp overlap. Actual start of this clone is at base position 1 of ZK75; actual end is at 22444 of ZK84.

NOTES:

Coding seqences below are the result of integration and manual review of the following data : computer analysis using the program Genefinder (P. Green and L. Hillier, personal communication), the large scale EST projects of Yuji Kohara (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C. elegans ORFeome cloning project (http://worfdb.dfci.harvard.edu/), similarity to other proteins from BlastX analyses (http://blast.wustl.edu/), sequence conservation with C. briggsae using Jim Kent's WABA alignment program (Genome Research 10:1115-1125, 2000), individual C. elegans GenBank submissions, and personal communications with C. elegans researchers. tRNAs are predicted using the program tRNAscan-SE (Lowe, T.M. and Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

```
FEATURES             Location/Qualifiers
     source          1. .17986
                     /organism="Caenorhabditis elegans"
                     /mol_type="genomic DNA"
                     /strain="Bristol N2"
                     /db_xref="taxon:6239"
                     /chromosome="II"
                     /clone="ZK75"
     gene            9642. .10286
                     /gene="ins-2"
                     /locus_tag="ZK75.2"
     CDS             join(9642. .9867,10192. .10286)
                     /gene="ins-2"
                     /locus_tag="ZK75.2"
                     /standard_name="ZK75.2"
                     /note="contains similarity to Pfam domain PF03488
                     (Nematode insulin-related peptide beta type); coded for by
                     the following C. elegans cDNAs: OSTR095A9_1"
                     /codon_start=1
                     /product="Insulin related protein 2"
                     /protein_id="AAC46745.1"
                     /db_xref="GI:733561"
                     /db_xref="WormBase:ZK75.2"
                     /translation="MNAIIFCLLFTTVTATYEVFGKGIEHRNEHLIINQLDIIPVEST
                     PTPNRASRVQKRLCGRRLILFMLATCGECDTDSSEDLSHICCIKQCDVQDIIRVCCPN
                     SFRK"
     gene            complement(11412. .11816)
                     /gene="ins-3"
                     /locus_tag="ZK75.3"
     CDS             complement(join(11412. .11497,11579. .11816))
                     /gene="ins-3"
                     /locus_tag="ZK75.3"
                     /standard_name="ZK75.3"
                     /note="contains similarity to Pfam domain PF03488
                     (Nematode insulin-related peptide beta type); coded for by
                     the following C. elegans cDNAs: OSTF095A10_1,
                     OSTF095A10_2, OSTR095A10_1"
                     /codon_start=1
                     /product="Insulin related protein 3"
                     /protein_id="AAC46746.1"
                     /db_xref="GI:733562"
                     /db_xref="WormBase:ZK75.3"
                     /translation="MKLSVVLALFIIFQLGAASLMRNWMFDFEKELEHDYDDSEIGFH
                     NIHSLMARSRRGDKVKICGTKVLKMVMVMCGGECSSTNENIATECCEKMCTMEDITTK
                     CCPSR"
     gene            17500. .17878
                     /gene="ins-4"
                     /locus_tag="ZK75.1"
     CDS             join(17500. .17734,17793. .17878)
                     /gene="ins-4"
                     /locus_tag="ZK75.1"
                     /standard_name="ZK75.1"
                     /note="contains similarity to Pfam domain PF03488
                     (Nematode insulin-related peptide beta type); coded for by
                     the following C. elegans cDNAs: OSTF095B8_1, OSTF113D9_1,
                     OSTR095B8_1, OSTR113D9_1"
                     /codon_start=1
                     /product="Insulin related protein 4"
                     /protein_id="AAC46744.1"
                     /db_xref="GI:733563"
                     /db_xref="WormBase:ZK75.1"
                     /translation="MFSFFTYFLLSALLLSASCRQPSMDTSKADRILREIEMETELEN
                     QLSRARRVPAGEVRACGRRLLLFVWSTCGEPCTPQEDMDIATVCCTTQCTPSYIKQAC
                     CPEK"
ORIGIN
```

```
Alignment Scores:
Pred. No.:              3.78e+03        Length:         17986
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-858-332G-28 (1-6) x U23451 (1-17986)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1258 CATAATCATAATCATAAT 1241
```

```
RESULT 388
AY277972/c
LOCUS       AY277972               18158 bp    DNA     linear   VRT 22-MAR-2004
DEFINITION  Takifugu rubripes transcription factor Sox6b (Sox6b) gene, complete
            cds.
ACCESSION   AY277972
VERSION     AY277972.1  GI:33415955
KEYWORDS    .
SOURCE      Takifugu rubripes (Fugu rubripes)
  ORGANISM  Takifugu rubripes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Takifugu.
REFERENCE   1  (bases 1 to 18158)
  AUTHORS   Koopman,P., Schepers,G., Brenner,S. and Venkatesh,B.
  TITLE     Origin and diversity of the Sox transcription factor gene family:
            genome-wide analysis in Fugu rubripes
  JOURNAL   Gene 328, 177-186 (2004)
   PUBMED   15019997
REFERENCE   2  (bases 1 to 18158)
  AUTHORS   Koopman,P., Schepers,G., Brenner,S. and Venkatesh,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-APR-2003) Institute of Molecular and Cell Biology,
            30, Medical Drive, Singapore 117609, Singapore
FEATURES             Location/Qualifiers
     source          1. .18158
                     /organism="Takifugu rubripes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:31033"
     gene            <101. .>16707
                     /gene="Sox6b"
     mRNA            join(<101. .295,2117. .2306,4181. .4270,5604. .5776,
                     6472. .6540,7666. .7786,8215. .8294,10215. .10349,
                     10972. .11134,12035. .12213,14179. .14290,14756. .14926,
                     15613. .15826,16383. .>16707)
                     /gene="Sox6b"
                     /product="transcription factor Sox6b"
     CDS             join(101. .295,2117. .2306,4181. .4270,5604. .5776,6472. .6540,
                     7666. .7786,8215. .8294,10215. .10349,10972. .11134,
                     12035. .12213,14179. .14290,14756. .14926,15613. .15826,
                     16383. .16707)
                     /gene="Sox6b"
                     /codon_start=1
                     /product="transcription factor Sox6b"
                     /protein_id="AAQ18515.1"
```

```
                     /db_xref="GI:33415956"
                     /translation="MSSVRTTLPFLPPSEEGITAGQDNLPWTNELKEAPSPLSHDQHR
                     TEEVQPIREKLANEWEEPAEVDNDSKKMSTYSSTSPSKPDESCRDPGDPVIHFGTPER
                     RKGSLADVVDSLKQKKLVELTKTEQEESDCVERLLSKDWRERVDRLNATELLGDVKGT
                     PESLEEKEQQLSNMIGQLIGLREQLLSAHDEQKKMATSQLEKQRQQMELARQQQEQIA
                     RQQQQLLQQQHKINLLQQQIQVQGHMTPMMIPVFPHDQRSLVAATATQQGFLFPQRMS
                     YKPGDNYPMQFIPSTMAAAAAAGLSPLQLQQLYAAQLASMQISQGTKMKPPSFSPLSP
                     STLKSEKRPASPVGHIKEEVSTQPLNLSARSQTAETSPTHSFISGNKSSPTPLVKGRV
                     PSPAPNKGRNTSLDILSSLNSTALFGDQDAVMKAIQEARSMREQIQREQLNQQPSLEA
                     KLSALSNMSLSNGSKQERLHYEALSQHLGKLGEDAGKMVHRVIDLTRPEDLEGNDISE
                     ARVFRESRGRNGNEPHIKRPMNAFMVWAKDERRKILQTFPDMHNSNISKILGSRWKAM
                     TNQEKQPYYEEQARLSKIHLEKYPNYKYKPRPKRTCIIDGKKLRIGEYKQLMRSRRQE
                     MRQFFVGPPSSVPLANSPSGLPVYPGAITMAKAATQSPQLTSDCSSNSASPEPAIPVI
                     HSAYRIKSETGGRSSGMAGLDPPSDRTIGDDTDMYDEFDEEPKSDYSSDHETIGAN"
ORIGIN

Alignment Scores:
Pred. No.:                3.82e+03        Length:         18158
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       5               Gaps:           0

US-09-858-332G-28 (1-6) x AY277972 (1-18158)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      17170 CACAACCACAACCACAAC 17153

RESULT 389
AY520825
LOCUS       AY520825               20247 bp    DNA     linear   ROD 01-MAR-2005
DEFINITION  Mus musculus ISP1 (Isp1) and ISP2 (Isp2) genes, complete cds.
ACCESSION   AY520825
VERSION     AY520825.1  GI:49487500
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 20247)
  AUTHORS   O'Sullivan,C.M., Tang,L., Xu,H. and Rancourt,D.E.
  TITLE     Origin of the Implanation Serin Proteinase Subfamily in Mouse
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 20247)
  AUTHORS   O'Sullivan,C.M., Tang,L., Xu,H. and Rancourt,D.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JAN-2004) Biochemistry & Molecular Biology, U of
            Calgary, 3330 Hospital Dr. N.W., Calgary, AB T2J 1N4, Canada
FEATURES             Location/Qualifiers
     source          1. .20247
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="129/SvJ"
                     /db_xref="taxon:10090"
                     /chromosome="17"
                     /map="A3.3"
     gene            4730. .7968
                     /gene="Isp1"
     mRNA            join(4730. .4891,5496. .5556,5738. .5918,6020. .6285,
                     7321. .7484,7610. .7968)
                     /gene="Isp1"
                     /product="ISP1"
     5'UTR           4730. .4891
                     /gene="Isp1"
     CDS             join(5496. .5556,5738. .5918,6020. .6285,7321. .7484,
                     7610. .7762)
                     /gene="Isp1"
                     /note="serine proteinase"
                     /codon_start=1
                     /product="ISP1"
                     /protein_id="AAT66743.1"
                     /db_xref="GI:49487501"
                     /translation="MFRLLLLALSCLESTVFMASVSISRSKPVGIVGGQRTPPGKWPW
                     QVSLRMYSYEVNSWVHICGGSIIHPQWILTAAHCIQSQDADPAVYRVQVGEVYLYKEQ
                     ELLNISRIIIHPDYNDVSKRFDLALMQLTALLVTSTNVSPVSLPKDSSTFDSTDQCWL
                     VGWGNLLQRVPLQPPYQLHEVKIPIQDNKSCKRAYRKKSSDEHKAVAIFDDMLCAGTS
                     GRGPCFGDSGGPLVCWKSNKWIQVGVVSKGIDCSNNLPSIFSRVQSSLAWIHQHIQ"
     3'UTR           7763. .7968
                     /gene="Isp1"
     misc_feature    7969. .14673
                     /note="intergenic region"
     gene            14674. .18578
                     /gene="Isp2"
     mRNA            join(14674. .14836,16305. .16365,16518. .16698,16894. .17159,
                     17990. .18156,18241. .18578)
                     /gene="Isp2"
                     /product="ISP2"
     5'UTR           14674. .14836
                     /gene="Isp2"
     CDS             join(16305. .16365,16518. .16698,16894. .17159,17990. .18156,
                     18241. .18405)
                     /gene="Isp2"
                     /note="serine proteinase"
                     /codon_start=1
                     /product="ISP2"
                     /protein_id="AAT66744.1"
                     /db_xref="GI:49487502"
                     /translation="MLIQLCLTLFFLGCSIAGTPAPGPEDVLMGIVGGHSAPQGKWPW
                     QVSLRIYRYYWAFWVHNCGGSIIHPQWVLTAAHCIRERDADPSVFRIRVGEAYLYGGK
                     ELLSVSRVIIHPDFVHAGLGSDVALLQLAVSVQSFPNVKPVKLPSESLEVTKKDVCWV
                     TGWGAVSTHRSLPPPYRLQQVQVKIIDNSLCEEMYHNATRHRNRGQKLILKDMLCAGN
                     QGQDSCYGDSGGPLVCNVTGSWTLVGVVSWGYGCALRDFPGVYARVQSFLPWITQQMQ
                     RFS"
     3'UTR           18406. .18578
                     /gene="Isp2"
ORIGIN

Alignment Scores:
Pred. No.:                4.24e+03        Length:         20247
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       9               Gaps:           0

US-09-858-332G-28 (1-6) x AY520825 (1-20247)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      12772 CACAACCACAACCACAAC 12789

RESULT 390
AE002104
LOCUS       AE002104               21206 bp    DNA     linear   BCT 04-JUN-2004
DEFINITION  Ureaplasma parvum serovar 3 str. ATCC 700970 section 5 of 59 of the
            complete genome.
ACCESSION   AE002104 AF222894
VERSION     AE002104.1  GI:6898990
KEYWORDS    .
SOURCE      Ureaplasma parvum serovar 3 str. ATCC 700970
  ORGANISM  Ureaplasma parvum serovar 3 str. ATCC 700970
            Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Ureaplasma.
REFERENCE   1  (bases 1 to 21206)
  AUTHORS   Glass,J.I., Lefkowitz,E.J., Glass,J.S., Heiner,C.R., Chen,E.Y. and
            Cassell,G.H.
  TITLE     The complete sequence of the mucosal pathogen Ureaplasma
            urealyticum
  JOURNAL   Nature 407 (6805), 757-762 (2000)
   PUBMED   11048724
REFERENCE   2  (bases 1 to 21206)
  AUTHORS   Glass,J.I., Lefkowitz,E.J., Glass,J.S., Heiner,C.R., Nguyen,T.,
            Chen,E.Y. and Cassell,G.H.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (10-JAN-2000) Microbiology, University of Alabama at
            Birmingham, BBRB 276/11, Birmingham, AL 35294-2170, USA
FEATURES             Location/Qualifiers
     source          1. .21206
                     /organism="Ureaplasma parvum serovar 3 str. ATCC 700970"
                     /mol_type="genomic DNA"
                     /strain="ATCC 700970"
                     /serovar="3"
                     /db_xref="ATCC:700970"
                     /db_xref="taxon:273119"
                     /note="biovar: 1"
     gene            114. .2630
                     /locus_tag="UU043"
     CDS             114. .2630
                     /locus_tag="UU043"
                     /codon_start=1
                     /transl_table=4
                     /product="conserved hypothetical membrane lipoprotein"
                     /protein_id="AAF30448.1"
                     /db_xref="GI:6898991"
                     /translation="MKKLKLITLSLSPIPIVAIMATACSTTKKEKEVDHKQDELTTIN
                     NNLQLEYPQNKDIETSAAIKEEIKPLNLPSGVVFSIKEISPKANDESTLIVKYTLKKG
                     NITKEFKKEINGFKKIRKDQNIANLKDMHDDFKTIFNNIKLKDTFDFKLKGLFNSNFG
                     EYDKLLPSQIINNFAQGVETRINTNQDRLGIEVIDVSYPNKNFGSANREGNLKMSLLV
                     TDKKTKQVFIKSIIAYGFKTNAMGLDENGSIPGGGADLVKPKVDENNYFNKTQLERYE
                     IDNQAYLNGLKGQHMGKTWQQVRPELGDNEQKINEFDQKAKGVSQDAYASAAYKGFTL
                     PVYDKNGDFKGLSINESSYGQARSWVDTRGRDEWKTTGLPRTLPNEKYRDEALQTLGI
                     SSLTLKDGKKDTWDKSSGTTWILDYQKTNDNKYPTKWYFATNLHVADTITDKTTSIDL
                     MKLMDSVKTKTTLRLSNLDENIYRFGFASKNNKFLLNHGLKKIYDGRNFLKTKPTEYL
                     TEQQKEKYKDAGSFVDFAVFELDFEQLKLLNVWNNQLNSNGGVITKYDHSSAQDLAKI
                     ITSDYANHSDKQIKFLSKSYLNDYSKIDVPLRSTSYKFEGKDELFALGWPNSTRDGFF
                     ERYVDDDQTKYRTTDNFSLWTNSDYRFFGKLTEQEGGQPAFPTERTERGNYLSYAIGY
                     RSFIDKPGVLDAFISAPHTGNDLYKSLDGKKYINMGLEYMPRHYAPAGGASGSSVRNQ
                     NNEIVGIYHVSNEFASTGLATAFRSEGYDYKGLYGDKYKELPQYDLIYGGGKDQATGQ
                     SYREKMKEIYKNNNNIKTALFPNGFDQYDEKFKFNNNQSK"
     gene            2653. .5001
                     /locus_tag="UU044"
     CDS             2653. .5001
                     /locus_tag="UU044"
                     /codon_start=1
                     /transl_table=4
                     /product="unique hypothetical"
                     /protein_id="AAF30449.1"
                     /db_xref="GI:6898992"
                     /translation="MKFIKRKTKLLTITIGAVAVSSILLGGIFYGTSQKSPSSFGIAS
                     IDQKENFINKDNLDYQKARPSIKDSNLKEIPKPKPQPKPKPQPTPFPDPIPTPPKKEE
                     LKKPDIKPEEPKKPEIKPEPKPEPIPQPAPPIETKPKEELLPPNPPPPKEEPKPEPNP
                     QPQPQQVPNNSNSRIIEINGVRVEAEVEVPPPRDIAEYDKQNNLVNPNPYINDSVGKI
                     KNVKVTDELRKATGKLVQGNLGRWDYKHLINDLLTLKPEEIEKYVKNDKSGYYAKVWY
                     RFSRLFESENVVNFLTEQGKKEYPEMKSKFVSKDHKYAWLYQHLDLTKFTQLSNESES
                     YLKEGYTPDPDNAYVDENGKISSHAYSPAKGYNSVTSRMENDNWNRRVFGYKSWYGRT
                     PGNLVEGNYPGWKKTNVTQEFHQYGVSDGDGITVNKLTREKTEDGRLNEGYVIDIDAD
                     NPQGYEKTKKLIQTLKEKNINITGYRIHNMGKSDSSQKFVDILKTLPNQLPLLELFFS
                     AGSHNTSSLIALKDKKIKELGLFTLGNSLLDEWSINPNALRNVEWINSNDYNVSFNYK
                     QGADIATRITFDTLAFDESDYNDNASDIKSKLKQINDGLRMVYWTRNNEPIFQGSFGP
                     GLDPDHKESGNSYPQGLDFSRVPQIRSLRGLIFKDEQKTSNNKDRKLRRINFYNNSTT
                     YKMSIEDLNEAGFNEHIVSGEPGEKSKITFSNGSGTTKIQIDGDQELSANGISNLSAM
                     FNFAESLQRTIVVNNTNSPLANQLRNAGYSVESTTNAGLIDI"
     gene            5032. .7536
                     /locus_tag="UU045"
     CDS             5032. .7536
                     /locus_tag="UU045"
                     /codon_start=1
                     /transl_table=4
                     /product="conserved hypothetical membrane lipoprotein"
                     /protein_id="AAF30450.1"
                     /db_xref="GI:6898993"
                     /translation="MKKNKLTTLALILPITILTPIVIASCTNKTKVKKSSSLDKIASN
                     LKLEYFNNKANTKASSVQKDEIKKPLNLPNDVVFSVKDVFVSHKDQSVLIVKYTLKKG
                     NEIQEYTYEIKGFKSVYEKDKIVNDLSQANEDFKKIVNNIRLKDTFDFKLAAFPNQNY
                     DQLLPSQIYKNYYQGIEIQQHKYQNELDIKIINFLYPDGDFGSANKNGTLKLSLMLTD
                     KKNNQVYYKLLEVSGFKSNPYGVDENGTIPGIGTERLKPKNQDDYFSKTQLQRYEIDN
                     EGYLQILKRQNNDKNWKELRPDLNATVSDIKHFDEKAKNVGQDSYESAAYKGFTLPVY
                     ESDGKISGLALAGKDTPKGPSWVDAIGRNQWQIGGLPRTLPNEKYRQEAMQTFSLGIL
                     NNDSHKNNTYNKTAGTTWILDYQKTSDNKYPTKWYFATNLHVADAINENTLSINLMRL
                     MDSAQIKTTFRLSNLDENIYNFGFRSKEHGKNLLNHGLKKIFDGRDFLKTKPAEYLIN
                     SQKEKYKDVGNFTDFAVFELDFEKLELVNVWKNFLGENNGLVTKYNNYNPQELAKVIT
                     SNYANNKNNQIKFLSKSYLSDYSKIDVPLKYRQEDAKTWFKKYDELFALGWPNSTEDF
                     FFKAYVDDDQLKYRTRDNFSLWTNSDYRFFNNLTQQEGGQPAFPPERTERGNYLSYAI
                     GFRSFIQKPGIVDAFIAVPQIGNNLYTSSDNKKYINMGLEYLPKHFAPAGGASGTSVR
                     NQKNELVAIYHAKYDSSKTGLAAAFRSEGYDYQGLYGNYNLPQYDLIYGGGKDQTEKK
                     SYREAMKDIYQNNNIKTALFPDGFDKIPDEFKFNNN"
     gene            7573. .9948
                     /locus_tag="UU046"
     CDS             7573. .9948
                     /locus_tag="UU046"
                     /codon_start=1
                     /transl_table=4
                     /product="unique hypothetical"
                     /protein_id="AAF30451.1"
                     /db_xref="GI:6898994"
                     /translation="MKFIKRKTKLLTITIGAVAVSSILLGGIFYGTSQKSPSSFGIAS
                     IDQKENFINKDNLDYQKARPSIKDNNLKEIPKPKPQPKPEPQPTPFPDPIPTPPKKEE
                     LKKPEIKPEEPKKPEIKPEPIPKPKPQPIPQPTPPVETKPKEELLPPSPPPPKEEPKP
                     EPDPQPQPQQIPNQSTVRKIELNGVLVDAEVEVPPPRQTFKYDQDNGLSNLNPYTNIS
                     VGKIKKVFVTDELRHKSADLVRGNLKRGDYQSLVKDLLDPNIKPEEIDSYIAMVDKSG
                     YHAKLWSKFKKLFDTDNVVNFLNEQGKKEYPNMKTKFVSDAHKYAWLYAHLDFSKFTK
                     LSANSEKYLQEGLTPDPDNSYVNENGELDSYAYSPAKEYNTVTSRLANDNANRRVFGY
                     NEWYNRSPNGLANGDYPGWNKSDATAEFKQYGIKDGDGIKVYKLERQKPQEGKLNTGY
                     IVDIDADNPDGYQKTKELIQKLNNQNKKITGYRIRNMGKSDSGQKFSDILKALPNELP
                     LLELFFSAGSHNTSALSALETKHIKELGLYTLGNSLLDEWSINPNALRKVEWINSNDY
                     NVSSEYKQGSDIATRITFDTLSFDKNDFNDNANDLRTKLKRINDGLRMVYWTRNNEPF
                     FQGGFGPGLDPDHKEVGNSYPQGLDFSRVPQIRSLRGLIFKDEQKASNNRERKLRRVN
                     FFNDKENYEMSINDLNEAGFSEHIVTNEPMPPKSKITFSNGNATKRIYIKGNGSLTAS
                     GIQNLATLFNLAESLDSKSVVVDSNNSELKSQLEGLGYKVSDASDANYIDI"
     gene            9963. .12566
                     /locus_tag="UU047"
     CDS             9963. .12566
                     /locus_tag="UU047"
                     /codon_start=1
                     /transl_table=4
                     /product="conserved hypothetical ATP/GTP-binding protein"
                     /protein_id="AAF30452.1"
                     /db_xref="GI:6898995"
                     /translation="MKIKKLIASVSIITPIVFVSALAASCVNNKHEDKNIVHNSSDHG
                     NNKKNNFTNDLNQNLTTHEFKSNLISTSQLDNIAKLINFTYENKANTYLKNILINQLK
                     HSPIQNQDFKLEIIGLYPKQNSLQDLIIYYKLTNQKTKEIKGYYFELNGFKKPDNTIF
                     SNQLTPELKKIIDTIQFKKTFDLVINNQQVLNYENVLPSQIKNQLLMGLKVIQKEYTD
                     KIKLSVLDVLFINEGGQINANKLGAFSLLLEVLDLKSKKTYQILIPVDKFKTNPYGAD
                     EYDLLPTQVNNGFAPTSLAQINQYNNADQQTRYLYDNENYLKSLKAYQRNVNWYQIRQ
                     DLINNKQKINEFDQKAPQVFQDSYESAARKGFTLPVYDEDGKYQGLSFNETEIGKSVS
                     WVDAIGKDQWKINGLARTLPNDMYKQIALQTFGIQIQTPNGKPRESDVVAGTMWIMDY
                     QKRNDNKYPTKWYFGTNLHVAEALKSTTTVFGINKIMPTVKTKTTLGLANADDNIYRF
                     SLVSKDNQPNMDKPISNGIKTIYDGRDFLKLNPSDLLTSKLKNKYHNLQEFVDFAVFE
                     IDFEKIKLGSITKNFYSGSDWSVDKYNNLDPSELAKLITNDYAQKPNEQIKFLSKSYL
                     NDYARINVPLDSKYNQAFDTNKYDELYAVGWPSSASDYFLDPIKDQKQFESRRESYSL
                     WINSNYQFYNKLNHNPPLFPLTQINRGDFLSYNIGYRSFTNKPGLLDAFIASPITGKT
                     VHKSTDNQNYIGFGLNYLPMHYSPIGGSSGTSLRNQKNELVGIWHVGNGFAQTGLAAA
                     FRSEGYDYHGLYGSYNLPQYDLIYGGGKDQKTSYRQAMMALYKNKNIATALFPNGFGE
                     DQIPSAFKFKN"
     gene            12596. .14086
                     /locus_tag="UU048"
     CDS             12596. .14086
                     /locus_tag="UU048"
                     /codon_start=1
                     /transl_table=4
                     /product="unique hypothetical"
                     /protein_id="AAF30453.1"
                     /db_xref="GI:6898996"
                     /translation="MMNTKTFTSVNRVIYDDNYSLKQQQKSSFINQFLKGLLSAITLL
                     FFILLLIFAENTLFGLGFGDENKSMMISKSLNAFFDLHSPKYLQLNFLIVFRFFILSF
                     TLFYTLIKNFTNLYWHRATIKKYLPWFVLYLVIATISFLLFFTFFSVWPKEVFNLVFL
                     LLVLFLLNLSYEIFNYFISKKTNPLLYDNYKNLIIAMVFQALLLLFVIITPLVWINTG
                     KSPNFLFVDNRFYTRIVDIFTVQSGKNFIILIAFFFFLITFIVLANTNFFALVINKRY
                     DRNYVKNNLWFILLLFSAIFIWLLRVFAYKHENENLPVGNNHLLWVYILQSFFAIIIL
                     ILYMVFTLKKRLSVKSSLNTLLNLVVTQTILSLSLFLVTLFNSKSVVSLINVFITITV
```

```
                QMSVFGIYIFQNKNISTKLLVLLKVIMILIILTAAIVGFDYLLTSDHHNNYLFSNIQP
                KMNLVQIMLLLNFSLNFTLISYLTIKFAMVIFKINKLNKELNNEKK"
     gene       14073. .14588
                /locus_tag="UU049"
     CDS        14073. .14588
                /locus_tag="UU049"
                /codon_start=1
                /transl_table=4
                /product="unique hypothetical"
                /protein_id="AAF30454.1"
                /db_xref="GI:6898997"
                /translation="MKKNKSNLTPTTNYFDVFNTYKEKKASVDLITYEELMASVLFDN
                KLGFESEVYLDFVKKFTLAFEKKLDIWFENFIINFNLNLKFSTTIMIPILVTKANSTT
                DAINFRNDQNPVYNNFLISYNQKIKKLLLQNHPVQILPHLILFKSNLNGSLVLVFSEK
                IIASIEQKSGN"
     gene       14591. .15505
                /locus_tag="UU050"
     CDS        14591. .15505
                /locus_tag="UU050"
                /codon_start=1
                /transl_table=4
                /product="unique hypothetical"
                /protein_id="AAF30455.1"
                /db_xref="GI:6898998"
                /translation="MQLKDLINKKKNLNNINLKVSNERNIFLINIMKLNQRLTFFSKN
                AFEIKESILSLKRIYNIKHDMLRHEERKIFKFLNKINDRVLWIYLTEEQKYSTDSYSR
                YEQKILKTIKSNRDDFILIGQGAIEFGKNHNLNVLQTFNDSNIKNLTTQLTKMIMILY
                TFDNYKKVNFVINSNKNYDGHFTILPMNEFSFDKFINLQKCDSNIIDFQKVKIYPNLN
                EFINVQINVFLVNIINTLITESSFYKTKNGLVATNNILKELDDNLSKIQRKITRVKTE
                LQIEEINLLARQNMNEDDNDNDGGVYES"
     gene       15462. .15974
                /locus_tag="UU051"
     CDS        15462. .15974
                /locus_tag="UU051"

Alignment Scores:
Pred. No.:              4.43e+03        Length:         21206
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AE002104 (1-21206)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       8063 CACAACCACAACCACAAC 8080

RESULT 391
BOVCASAS2X/c
LOCUS       BOVCASAS2X              21246 bp    DNA     linear   MAM 26-APR-1993
DEFINITION  Bovine alpha s2 casein type A protein (CASAS2) gene, exons 1-18.
ACCESSION   M94327
VERSION     M94327.1  GI:162795
KEYWORDS    alpha s2 casein protein type A.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 21246)
  AUTHORS   Groenen,M.A., Dijkhof,R.J., Verstege,A.J. and van der Poel,J.J.
  TITLE     The complete sequence of the gene encoding bovine alpha s2-casein
  JOURNAL   Gene 123 (2), 187-193 (1993)
   PUBMED   8428658
REFERENCE   2  (bases 1 to 21246)
  AUTHORS   Groenen,M.A.M., Dijkhof,R.J.M., van der Poel,J.J., van Diggelen,R.
            and Verstege,E.J.M.
  TITLE     Multiple octamer binding sites in the promoter region of the bovine
            alpha s2 casein gene
  JOURNAL   Unpublished (1993)
COMMENT     Original source text: Bos taurus (strain Dutch Black and White)
            (library: EMBL3) female adult blood DNA.
FEATURES             Location/Qualifiers
     source          1. .21246
                     /organism="Bos taurus"
                     /mol_type="genomic DNA"
                     /strain="Dutch Black and White"
                     /db_xref="taxon:9913"
                     /sex="female"
                     /tissue_type="blood"
                     /dev_stage="adult"
                     /germline
                     /tissue_lib="EMBL3"
     satellite       26. .227
                     /gene="CASAS2"
                     /standard_name="BDF element"
                     /note="putative"
                     /rpt_family="qlu-like artiodactyla retroposons"
                     /rpt_type=dispersed
     satellite       1681. .1882
                     /gene="CASAS2"
                     /standard_name="BMF element"
                     /note="putative"
                     /rpt_family="alu-like artiodactyla retroposon"
                     /rpt_type=dispersed
                     /label=repeat
     protein_bind    2030. .2039
                     /gene="CASAS2"
                     /standard_name="octamer binding protein Oct-1"
                     /note="binding shown with gelretardation compitition
                     experiments"
                     /citation=[2]
                     /bound_moiety="Oct-1"
                     /function="possibly initiation of transcription (not
                     proven)"
                     /evidence=experimental
                     /label=Oct-1
     protein_bind    2243. .2253
                     /gene="CASAS2"
                     /standard_name="Octamer binding protein Oct-1"
                     /note="binding shown with gelretardation experiments"
                     /citation=[2]
                     /bound_moiety="Oct-1"
                     /function="initiation of transcription (not proven)"
                     /evidence=experimental
     protein_bind    2292. .2302
                     /gene="CASAS2"
                     /standard_name="Octamer binding protein Oct-1"
                     /note="binding shown with gelretardation competition
                     experiments"
                     /citation=[2]
                     /bound_moiety="Oct-1"
                     /function="possibly initiation of transcription (not
                     proven)"
                     /evidence=experimental
     protein_bind    2413. .2423
                     /gene="CASAS2"
                     /standard_name="mammary gland factor"
                     /note="transcription factor MGF is mammary gland specific
                     but not lactation specific. See also Schmitt-Ney et al.
                     (1991) Mol.Cell. Biol. 11, 3745-3755"
                     /citation=[2]
                     /bound_moiety="MGF"
                     /function="initiation of transcription"
                     /evidence=experimental
     protein_bind    2455. .2464
                     /gene="CASAS2"
                     /standard_name="Octamer binding protein Oct-1"
                     /note="binding shown with gelretardation experiments"
                     /citation=[2]
                     /bound_moiety="Oct-1"
                     /function="initiation of transcription (not proven)"
                     /evidence=experimental
     TATA_signal     2480. .2486
```

```
                /gene="CASAS2"
                /note="putative"
     gene       join(2510. .2553,4339. .4401,6211. .6237,7027. .7051,
                7136. .7177,7567. .7593,8386. .8412,8853. .8879,9922. .9945,
                10042. .10086,10347. .10469,10999. .11025,12412. .12438,
                14134. .14157,14249. .14293,15398. .15517,16274. .16318,
                20705. .20970)
                /gene="CASAS2"
     mRNA       join(2510. .2553,4339. .4401,6211. .6237,7027. .7051,
                7136. .7177,7567. .7593,8386. .8412,8853. .8879,9922. .9945,
                10042. .10086,10347. .10469,10999. .11025,12412. .12438,
                14134. .14157,14249. .14293,15398. .15517,16274. .16318,
                20705. .20970)
                /gene="CASAS2"
     exon       2510. .2553
                /gene="CASAS2"
                /number=1
                /evidence=experimental
     intron     2554. .4338
                /gene="CASAS2"
                /number=1
                /evidence=experimental
     exon       3075. .3118
                /gene="CASAS2"
                /note="extra exon found in 4% of the ovine alpha s2
                casein. Conserved in the bovine gene; putative"
                /citation=[1]
                /function="unknown, probably none"
     repeat_region 3716. .3761
                /note="putative"
                /rpt_type=direct
     variation  3985. .3990
                /gene="CASAS2"
                /note="in the gene coding for the alpha s2 casein type A,
                the KpnI site at this position as presented in this entry
                is not always present"
                /citation=[1]
                /label=RFLP
     exon       4339. .4401
                /gene="CASAS2"
                /number=2
                /evidence=experimental
     intron     4402. .6210
                /gene="CASAS2"
                /number=2
                /evidence=experimental
     satellite  4894. .5086
                /gene="CASAS2"
                /standard_name="BMF element"
                /note="putative"
                /rpt_family="Alu-like artiodactyla retroposon"
                /rpt_type=dispersed
     satellite  5349. .5510
                /gene="CASAS2"
                /standard_name="BMF element"
                /note="putative"
                /rpt_family="Alu-like artiodactyla retroposon"
                /rpt_type=dispersed
     exon       6211. .6237
                /gene="CASAS2"
                /number=3
                /evidence=experimental
     intron     6238. .7026
                /gene="CASAS2"
                /number=3
                /evidence=experimental
     exon       7027. .7051
                /gene="CASAS2"
                /number=4
                /evidence=experimental
     intron     7052. .7135
                /gene="CASAS2"
                /number=4
                /evidence=experimental
     exon       7136. .7177
                /gene="CASAS2"
                /number=5
                /evidence=experimental
     intron     7178. .7566
                /gene="CASAS2"
                /number=5
                /evidence=experimental
     exon       7567. .7593
                /gene="CASAS2"
                /number=6
                /evidence=experimental
     intron     7594. .8385
                /gene="CASAS2"
                /number=6
                /evidence=experimental
     exon       8386. .8412
                /gene="CASAS2"
                /number=7
                /evidence=experimental
     intron     8413. .8852
                /gene="CASAS2"
                /number=7
                /evidence=experimental
     exon       8853. .8879
                /gene="CASAS2"
                /number=8
                /evidence=experimental
     intron     8880. .9921
                /gene="CASAS2"
                /number=8

Alignment Scores:
Pred. No.:              4.44e+03        Length:         21246
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     4               Gaps:           0

US-09-858-332G-28 (1-6) x BOVCASAS2X (1-21246)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       7288 CACAATCACAATCATAAT 7271

RESULT 392
AC017390
LOCUS       AC017390               21693 bp    DNA     linear   HTG 09-DEC-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC017390
VERSION     AC017390.1  GI:6553596
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 21693)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10210545 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .21693
                     /organism="Drosophila melanogaster"
```

```
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
ORIGIN

Alignment Scores:
Pred. No.:                4.53e+03        Length:         21693
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       14              Gaps:           0

US-09-858-332G-28 (1-6) x AC017390 (1-21693)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       4580 CACAACCATAACCACAAC 4597

RESULT 393
AC006656
LOCUS       AC006656               22683 bp    DNA     linear   INV 21-SEP-2004
DEFINITION  Caenorhabditis elegans fosmid H12I13, complete sequence.
ACCESSION   AC006656
VERSION     AC006656.1  GI:4263282
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 22683)
  AUTHORS   .
  CONSRTM   WormBase Consortium
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
  JOURNAL   Science 282 (5396), 2012-2018 (1998)
   PUBMED   9851916
REFERENCE   2  (bases 1 to 22683)
  AUTHORS   Ozersky,P. and Schaller,M.
  TITLE     The sequence of C. elegans fosmid H12I13
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 22683)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 22683)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 22683)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 22683)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   7  (bases 1 to 22683)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   8  (bases 1 to 22683)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2001) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   9  (bases 1 to 22683)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAY-2002) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   10 (bases 1 to 22683)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2002) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   11 (bases 1 to 22683)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-NOV-2003) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   12 (bases 1 to 22683)
  AUTHORS   .
  CONSRTM   WormBase Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (21-SEP-2004) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
COMMENT     Submitted by:
                  Genome Sequencing Center
                  Department of Genetics, Washington University
                  St. Louis , MO  63110, USA, and
                  Sanger Centre, Hinxton Hall
                  Cambridge CB10 IRQ, England
                  email: submissions@watson.wustl.edu and jes@sanger.ac.uk

            NOTICE:  This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate chemistry
            or covered by high quality data (i.e., phred quality >= 30); an
            attempt was made to resolve all sequencing problems, such as
            compressions and repeats; all regions were covered by sequence from
            more than one m13 subclone.

            For a graphical representation of this clone sequence and its
            analysis see:
            http://www.wormbase.org/db/seq/sequence?name=H12I13;class=Sequence

                        NEIGHBORING CLONE INFORMATION

            The 5' clone is R12C12, 1980 bp overlap; the 3' clone is F21H12,
            200 bp overlap. Actual start of this clone is at base position
            28566 of R12C12; actual end is at 10191 of F21H12.

            NOTES:

            Coding seqences below are the result of integration and manual
            review of the following data : computer analysis using the program
            Genefinder (P. Green and L. Hillier, personal communication), the
            large scale EST projects of Yuji Kohara
            (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C.
            elegans ORFeome cloning project (http://worfdb.dfci.harvard.edu/),
            similarity to other proteins from BlastX analyses
            (http://blast.wustl.edu/), sequence conservation with C. briggsae
            using Jim Kent's WABA alignment program (Genome Research
            10:1115-1125, 2000), individual C. elegans GenBank submissions,
            and personal communications with C. elegans researchers. tRNAs
            are predicted using the program tRNAscan-SE (Lowe, T.M. and
            Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).
FEATURES                Location/Qualifiers
     source             1. .22683
                        /organism="Caenorhabditis elegans"
```

```
                     /mol_type="genomic DNA"
                     /strain="Bristol N2"
                     /db_xref="taxon:6239"
                     /chromosome="II"
                     /clone="H12I13"
     gene            complement(2489. .4570)
                     /gene="H12I13.1"
                     /locus_tag="H12I13.1"
     CDS             complement(join(2489. .3153,3356. .3576,3619. .4300,
                     4360. .4570))
                     /gene="H12I13.1"
                     /locus_tag="H12I13.1"
                     /standard_name="H12I13.1"
                     /note="contains similarity to Pfam domain PF04435 (Domain
                     of unknown function (DUF545))"
                     /codon_start=1
                     /product="Hypothetical protein H12I13.1"
                     /protein_id="AAF39881.1"
                     /db_xref="GI:7206722"
                     /db_xref="WormBase:H12I13.1"
                     /translation="MPNAECRHNAGFWSAAYAEMPILKFTAHAESAVMPKRKMNGISA
                     YADMPKKIFGGICRYAVMPKNSECRTGDAKDRRWEKMDDVHVSDVSWEEVERHSYGTG
                     SDDTSSAVLIVYTKVKDATDDFTYDVTPNAHVVDVPGISLPPARTSISMETLSANPLH
                     GHVVDAPVFYHQQNHSSRTSISMESLNYPSPGDAHQTEEQLPENQLTKNRKTPGCYLN
                     QCIKFLHTKSENVGKPFVRTTFYANFKNVVYYKGTATTLGRGLQPLLHDAIISSEYKD
                     FRKCQMLFVMSIQITNQSFLYRLAKNSFSITLDRFYRIEVCEYQSKKFSGVHKCPGLV
                     KEIAIKPTSDLQTGQNLVAHPMSSRRKMIRDNKSSNKVCRTSLSGGAKKKRKERNARS
                     EVGDKILGITETKTVTSWNESGIELQRKIGFGAYGSAYSVKYGGKTVVMKLIGIRDNE
                     QLLFNELIITQQLGKLAKNRVCPNFLEYCGSHVMTDVPVGMRRDPKAIRHLAILMARG
                     GGVLDNWKFKDYRQCVSVFCQLVMSLKVVKDSINMVHRDIHLWNVLVSRTKTTCLDYS
                     MDGKVKLNSYGVVMHLIDFSKSLYGG"
     gene            complement(6235. .7369)
                     /gene="H12I13.2"
                     /locus_tag="H12I13.2"
     CDS             complement(join(6235. .6771,6821. .7192,7298. .7369))
                     /gene="H12I13.2"
                     /locus_tag="H12I13.2"
                     /standard_name="H12I13.2"
                     /note="contains similarity to Pfam domain PF00443
                     (Ubiquitin carboxyl-terminal hydrolase)"
                     /codon_start=1
                     /product="Hypothetical protein H12I13.2"
                     /protein_id="AAF39880.1"
                     /db_xref="GI:7206721"
                     /db_xref="WormBase:H12I13.2"
                     /translation="MSRDDDNIDHLMSKLDLFGSLELQKYLDGQTNCQLESSNLENTA
                     SPSDGEGEENSMEYSYSPRGLNLSTGIANNGSSCWLNCLIQIMFNTPGVEFSLHQFAT
                     KPFLERMRDTKDGNRNKVALMFLLRSVFTDLRFSAERAINITSIIKVVDQLSPHPVSN
                     QQQDASEMLQNIVIWLKDAVDAAVAFQRIHVRECHEDIRLMEQKLKIIKLQEDRFDRI
                     FASNMYEETLGLRVHIQKNSLRESFEKFFFANDAYSVDVVMVQIVRLAQDGRKLHDNF
                     LFPGNLVTESSATKEKCYFKLTSVMVHRGECAQSGHFYCFTLSKWANVNS"
     gene            complement(11616. .13481)
                     /gene="H12I13.3"
                     /locus_tag="H12I13.3"
     CDS             complement(join(11616. .12047,12093. .12461,12508. .12820,
                     12946. .13481))
                     /gene="H12I13.3"
                     /locus_tag="H12I13.3"
                     /standard_name="H12I13.3"
                     /note="contains similarity to Parvo-like virus
                     Nonstructural protein.; TR:Q9QMH2"
                     /codon_start=1
                     /product="Hypothetical protein H12I13.3"
                     /protein_id="AAF39882.1"
                     /db_xref="GI:7206723"
                     /db_xref="WormBase:H12I13.3"
                     /translation="MSRDELRYAVRRLNIISYQQMRTMKKQQLLDLMREHSNNLYNIF
                     PPENGAYNYIERPQVIREQKLGEIFPFQEAIFADYREFHNTVRNTTYNNERGSNIEKF
                     FDKMRNRVINDKKETNVKVGFNVEIEMIKPGTNETINFFIQLENSRLLQEEDRDDFYT
                     EKCRKIEDRIHRQELNGSGILMRGAVNDIPKVEKLNNLAINVFSNDTNSIIPVHISKK
                     STSVSENRIIDLFLIKNEDRMHYCLVKNINYLIGNESSNHSGLNSVCRYCFNSFRSKT
                     KYDNHLEIMLNLIKALIQALNHGLKLKKIHRILTFKESAWMKKYIMLNTNLRKQAKND
                     FEKDFFKLMNNSVFGKTMENIRNRVNIRLVTDDDSALKLIRKPNFKNSTIITEDLLSV
```

```
                     ELTKTKLVFDKPIYFHYDVMQKHYGDKIKLCYQDTDSLIYEVETEDIYEDMGNMKEWF

Alignment Scores:
Pred. No.:              4.73e+03        Length:         22683
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-858-332G-28 (1-6) x AC006656 (1-22683)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      10497 CATAATCATAATCATAAT 10514

RESULT 394
AC006656/c
LOCUS       AC006656               22683 bp    DNA     linear   INV 21-SEP-2004
DEFINITION  Caenorhabditis elegans fosmid H12I13, complete sequence.
ACCESSION   AC006656
VERSION     AC006656.1  GI:4263282
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 22683)
  AUTHORS   .
  CONSRTM   WormBase Consortium
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
  JOURNAL   Science 282 (5396), 2012-2018 (1998)
   PUBMED   9851916
REFERENCE   2  (bases 1 to 22683)
  AUTHORS   Ozersky,P. and Schaller,M.
  TITLE     The sequence of C. elegans fosmid H12I13
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 22683)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 22683)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 22683)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 22683)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   7  (bases 1 to 22683)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   8  (bases 1 to 22683)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2001) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   9  (bases 1 to 22683)
  AUTHORS   Waterston,R.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAY-2002) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   10 (bases 1 to 22683)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2002) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   11 (bases 1 to 22683)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-NOV-2003) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   12 (bases 1 to 22683)
  AUTHORS   .
  CONSRTM   WormBase Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (21-SEP-2004) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
COMMENT     Submitted by:
                  Genome Sequencing Center
                  Department of Genetics, Washington University
                  St. Louis , MO  63110, USA, and
                  Sanger Centre, Hinxton Hall
                  Cambridge CB10 IRQ, England
                  email: submissions@watson.wustl.edu and jes@sanger.ac.uk

            NOTICE:  This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate chemistry
            or covered by high quality data (i.e., phred quality >= 30); an
            attempt was made to resolve all sequencing problems, such as
            compressions and repeats; all regions were covered by sequence from
            more than one m13 subclone.

            For a graphical representation of this clone sequence and its
            analysis see:
            http://www.wormbase.org/db/seq/sequence?name=H12I13;class=Sequence

                        NEIGHBORING CLONE INFORMATION

            The 5' clone is R12C12, 1980 bp overlap; the 3' clone is F21H12,
            200 bp overlap. Actual start of this clone is at base position
            28566 of R12C12; actual end is at 10191 of F21H12.

            NOTES:

            Coding seqences below are the result of integration and manual
            review of the following data : computer analysis using the program
            Genefinder (P. Green and L. Hillier, personal communication), the
            large scale EST projects of Yuji Kohara
            (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C.
            elegans ORFeome cloning project (http://worfdb.dfci.harvard.edu/),
            similarity to other proteins from BlastX analyses
            (http://blast.wustl.edu/), sequence conservation with C. briggsae
            using Jim Kent's WABA alignment program (Genome Research
            10:1115-1125, 2000), individual C. elegans GenBank submissions,
            and personal communications with C. elegans researchers. tRNAs
            are predicted using the program tRNAscan-SE (Lowe, T.M. and
            Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).
FEATURES            Location/Qualifiers
     source         1. .22683
                    /organism="Caenorhabditis elegans"
                    /mol_type="genomic DNA"
                    /strain="Bristol N2"
                    /db_xref="taxon:6239"
                    /chromosome="II"
                    /clone="H12I13"
     gene           complement(2489. .4570)
                    /gene="H12I13.1"
                    /locus_tag="H12I13.1"
     CDS            complement(join(2489. .3153,3356. .3576,3619. .4300,
                    4360. .4570))
                    /gene="H12I13.1"
                    /locus_tag="H12I13.1"
                    /standard_name="H12I13.1"
                    /note="contains similarity to Pfam domain PF04435 (Domain
                    of unknown function (DUF545))"
                    /codon_start=1
                    /product="Hypothetical protein H12I13.1"
                    /protein_id="AAF39881.1"
                    /db_xref="GI:7206722"
                    /db_xref="WormBase:H12I13.1"
                    /translation="MPNAECRHNAGFWSAAYAEMPILKFTAHAESAVMPKRKMNGISA
                    YADMPKKIFGGICRYAVMPKNSECRTGDAKDRRWEKMDDVHVSDVSWEEVERHSYGTG
                    SDDTSSAVLIVYTKVKDATDDFTYDVTPNAHVVDVPGISLPPARTSISMETLSANPLH
                    GHVVDAPVFYHQQNHSSRTSISMESLNYPSPGDAHQTEEQLPENQLTKNRKTPGCYLN
                    QCIKFLHTKSENVGKPFVRTTFYANFKNVVYYKGTATTLGRGLQPLLHDAIISSEYKD
                    FRKCQMLFVMSIQITNQSFLYRLAKNSFSITLDRFYRIEVCEYQSKKFSGVHKCPGLV
                    KEIAIKPTSDLQTGQNLVAHPMSSRRKMIRDNKSSNKVCRTSLSGGAKKKRKERNARS
                    EVGDKILGITETKTVTSWNESGIELQRKIGFGAYGSAYSVKYGGKTVVMKLIGIRDNE
                    QLLFNELIITQQLGKLAKNRVCPNFLEYCGSHVMTDVPVGMRRDPKAIRHLAILMARG
                    GGVLDNWKFKDYRQCVSVFCQLVMSLKVVKDSINMVHRDIHLWNVLVSRTKTTCLDYS
                    MDGKVKLNSYGVVMHLIDFSKSLYGG"
     gene           complement(6235. .7369)
                    /gene="H12I13.2"
                    /locus_tag="H12I13.2"
     CDS            complement(join(6235. .6771,6821. .7192,7298. .7369))
                    /gene="H12I13.2"
                    /locus_tag="H12I13.2"
                    /standard_name="H12I13.2"
                    /note="contains similarity to Pfam domain PF00443
                    (Ubiquitin carboxyl-terminal hydrolase)"
                    /codon_start=1
                    /product="Hypothetical protein H12I13.2"
                    /protein_id="AAF39880.1"
                    /db_xref="GI:7206721"
                    /db_xref="WormBase:H12I13.2"
                    /translation="MSRDDDNIDHLMSKLDLFGSLELQKYLDGQTNCQLESSNLENTA
                    SPSDGEGEENSMEYSYSPRGLNLSTGIANNGSSCWLNCLIQIMFNTPGVEFSLHQFAT
                    KPFLERMRDTKDGNRNKVALMFLLRSVFTDLRFSAERAINITSIIKVVDQLSPHPVSN
                    QQQDASEMLQNIVIWLKDAVDAAVAFQRIHVRECHEDIRLMEQKLKIIKLQEDRFDRI
                    FASNMYEETLGLRVHIQKNSLRESFEKFFFANDAYSVDVVMVQIVRLAQDGRKLHDNF
                    LFPGNLVTESSATKEKCYFKLTSVMVHRGECAQSGHFYCFTLSKWANVNS"
     gene           complement(11616. .13481)
                    /gene="H12I13.3"
                    /locus_tag="H12I13.3"
     CDS            complement(join(11616. .12047,12093. .12461,12508. .12820,
                    12946. .13481))
                    /gene="H12I13.3"
                    /locus_tag="H12I13.3"
                    /standard_name="H12I13.3"
                    /note="contains similarity to Parvo-like virus
                    Nonstructural protein.; TR:Q9QMH2"
                    /codon_start=1
                    /product="Hypothetical protein H12I13.3"
                    /protein_id="AAF39882.1"
                    /db_xref="GI:7206723"
                    /db_xref="WormBase:H12I13.3"
                    /translation="MSRDELRYAVRRLNIISYQQMRTMKKQQLLDLMREHSNNLYNIF
                    PFENGAYNYIERPQVIREQKLGEIFPFQEAIFADYREFHNTVRNTTYNNERGSNIEKF
                    FDKMRNRVINDKKETNVKVGFNVEIEMIKPGTNETINFFIQLENSRLLQEEDRDDFYT
                    EKCRKIEDRIHRQELNGSGILMRGAVNDIPKVEKLNNLAINVFSNDTNSIIPVHISKK
                    STSVSENRIIDLFLIKNEDRMHYCLVKNINYLIGNESSNHSGLNSVCRYCFNSFRSKT
                    KYDNHLEIMLNLIKALIQALNHGLKLKKIHRILTFKESAWMKKYIMLNTNLRKQAKND
                    FEKDFFKLMNNSVFGKTMENIRNRVNIRLVTDDDSALKLIRKPNFKNSTIITEDLLSV
                    ELTKTKLVFDKPIYFHYDVMQKHYGDKIKLCYQDTDSLIYEVETEDIYEDMGNMKEWF
```

```
Alignment Scores:
Pred. No.:              4.73e+03        Length:         22683
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-858-332G-28 (1-6) x AC006656 (1-22683)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       9012 CATAATCATAATCATAAT 8995

RESULT 395
AC020515
LOCUS       AC020515               23241 bp    DNA     linear   HTG 03-JAN-2000
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC020515
VERSION     AC020515.1  GI:6664382
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 23241)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10213526 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .23241
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
ORIGIN

Alignment Scores:
Pred. No.:              4.84e+03        Length:         23241
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AC020515 (1-23241)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       3742 CATAACCATAACCATAAT 3759

RESULT 396
CQ614936/c
LOCUS       CQ614936               23732 bp    DNA     linear   PAT 02-FEB-2004
DEFINITION  Sequence 42694 from Patent WO0171042.
ACCESSION   CQ614936
VERSION     CQ614936.1  GI:41665270
KEYWORDS    .
SOURCE      Drosophila sp.
  ORGANISM  Drosophila sp.
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1
  AUTHORS   Venter,J.C., Adams,M., Li,P.W. and Myers,E.W.
  TITLE     Detection kits, such as nucleic acid arrays, for detecting the
            expression of 10,000 or more Drosophila genes and uses thereof
  JOURNAL   Patent: WO 0171042-A 42694 27-SEP-2001;
            PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1. .23732
                     /organism="Drosophila sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:7242"
ORIGIN

Alignment Scores:
Pred. No.:              4.94e+03        Length:         23732
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x CQ614936 (1-23732)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      22233 CACAATCACAACCACAAT 22216

RESULT 397
AC018178
LOCUS       AC018178               23775 bp    DNA     linear   HTG 09-DEC-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC018178
VERSION     AC018178.1  GI:6553013
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 23775)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10213689 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .23775
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
ORIGIN

Alignment Scores:
Pred. No.:              4.95e+03        Length:         23775
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AC018178 (1-23775)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      18890 CACAACCATAATCACAAT 18907

RESULT 398
AX537040
LOCUS       AX537040               24076 bp    DNA     linear   PAT 22-NOV-2002
DEFINITION  Sequence 641 from Patent WO02064766.
ACCESSION   AX537040
```

```
VERSION     AX537040.1  GI:25263476
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1
  AUTHORS   Contreras,R.H., Eberhardt,I., Luyten,W.H. and Reekmans,R.J.
  TITLE     Bax-responsive genes for drug target identification in yeast and
            fungi
  JOURNAL   Patent: WO 02064766-A 641 22-AUG-2002;
            JANSSEN PHARMACEUTICA N.V. (BE)
FEATURES             Location/Qualifiers
     source          1. .24076
                     /organism="Candida albicans"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:5476"
ORIGIN
```

Alignment Scores:

Pred. No.:	5.01e+03	Length:	24076
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-09-858-332G-28 (1-6) x AX537040 (1-24076)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      16874 CACAACCACAACCACAAC 16891
```

```
RESULT 399
AC014420/c
LOCUS       AC014420               25849 bp    DNA     linear   HTG 16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC014420
VERSION     AC014420.1  GI:6436915
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 25849)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10211482 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .25849
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
ORIGIN
```

Alignment Scores:

Pred. No.:	5.36e+03	Length:	25849
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	14	Gaps:	0

US-09-858-332G-28 (1-6) x AC014420 (1-25849)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      16987 CATAACCATAACCACAAC 16970
```

```
RESULT 400
AC013846
LOCUS       AC013846               26857 bp    DNA     linear   HTG 16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC013846
VERSION     AC013846.1  GI:6437489
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 26857)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10210846 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .26857
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
ORIGIN
```

Alignment Scores:

Pred. No.:	5.56e+03	Length:	26857
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	14	Gaps:	0

US-09-858-332G-28 (1-6) x AC013846 (1-26857)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      24335 CATAATCACAACCATAAT 24352
```

```
RESULT 401
AC019865/c
LOCUS       AC019865               28248 bp    DNA     linear   HTG 03-JAN-2000
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC019865
VERSION     AC019865.1  GI:6665032
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 28248)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10211213 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .28248
```

```
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
ORIGIN

Alignment Scores:
Pred. No.:                5.84e+03        Length:         28248
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       14              Gaps:           0

US-09-858-332G-28 (1-6) x AC019865 (1-28248)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      24033 CATAACCATAACCATAAT 24016

RESULT 402
CQ578696/c
LOCUS       CQ578696               28360 bp    DNA     linear   PAT 02-FEB-2004
DEFINITION  Sequence 6454 from Patent WO0171042.
ACCESSION   CQ578696
VERSION     CQ578696.1  GI:41641110
KEYWORDS    .
SOURCE      Drosophila sp.
  ORGANISM  Drosophila sp.
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1
  AUTHORS   Venter,J.C., Adams,M., Li,P.W. and Myers,E.W.
  TITLE     Detection kits, such as nucleic acid arrays, for detecting the
            expression of 10,000 or more Drosophila genes and uses thereof
  JOURNAL   Patent: WO 0171042-A 6454 27-SEP-2001;
            PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1. .28360
                     /organism="Drosophila sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:7242"
ORIGIN

Alignment Scores:
Pred. No.:                5.86e+03        Length:         28360
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x CQ578696 (1-28360)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      22869 CACAATCACAATCACAAT 22852

RESULT 403
AY218787/c
LOCUS       AY218787               28708 bp    DNA     circular SYN 21-APR-2003
DEFINITION  Cloning vector pAJ1, complete sequence.
ACCESSION   AY218787 AC138659
VERSION     AY218787.1  GI:30027630
KEYWORDS    .
SOURCE      Cloning vector pAJ1
  ORGANISM  Cloning vector pAJ1
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 28708)
  AUTHORS   McCombie,W.R., Martienssen,R., de la Bastide,M., Preston,R.,
            Spiegel,L., Nascimento,L., Zutavern,T., Balija,V., Bell,M.,
            Miller,B., Muller,S., Katzenberger,F., Yang,C., Dike,S.,
            O'Shaughnessy,A., Dedhia,N. and Palmer,L.
  TITLE     Complete sequence of cloning vector pAJ1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 28708)
  AUTHORS   McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JAN-2003) Lita Annenberg Hazen Genome Center, Cold
            Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor,
            NY 11724, USA
COMMENT     On Apr 21, 2003 this sequence version replaced gi:27753654.
FEATURES             Location/Qualifiers
     source          1. .28708
                     /organism="Cloning vector pAJ1"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:218927"
ORIGIN

Alignment Scores:
Pred. No.:                5.93e+03        Length:         28708
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       11              Gaps:           0

US-09-858-332G-28 (1-6) x AY218787 (1-28708)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      19313 CACAACCACAACCACAAC 19296

RESULT 404
YSCH9433/c
LOCUS       YSCH9433               29276 bp    DNA     linear   PLN 19-MAR-2002
DEFINITION  Saccharomyces cerevisiae chromosome VIII cosmid 9433.
ACCESSION   U11582 U00093
VERSION     U11582.1  GI:2289793
KEYWORDS    .
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 29276)
  AUTHORS   Favello,T.
  TITLE     The sequence of S. cerevisiae cosmid 9433
  JOURNAL   Unpublished (1994)
REFERENCE   2  (bases 1 to 29276)
  AUTHORS   Johnston,M., Andrews,S., Brinkman,R., Cooper,J., Ding,H., Dover,J.,
            Du,Z., Favello,A., Fulton,L., Gattung,S., Geisel,C., Kirsten,J.,
            Kucaba,T., Hillier,L., Jier,M., Johnston,L., Keppler,D.,
            Langston,Y., Latreille,P., Louis,E., Macri,C., Mardis,E.,
            Mouser,L., Nhan,M., Rifken,L., Riles,L., St.Peter,H., Thornton,L.,
            Trevaskis,E., Vaudin,M., Vaughan,K., Vignati,D., Wilcox,L.,
            Willis,A., Wilson,R., Wohldman,P. and Waterston,R.
  TITLE     The complete sequence of Saccharomyces cerevisiae chromosome VIII
  JOURNAL   Unpublished (1994)
REFERENCE   3  (bases 1 to 29276)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUN-1994) Washington University, St. Louis, MO 63110,
            USA
COMMENT     On Aug 1, 1997 this sequence version replaced gi:2289791.
            Submitted by:
                     Genome Sequencing Center
                     Department of Genetics, Washington University,
                     St. Louis, MO  63110, USA
                     e-mail: mj@sequencer.wustl.edu.
FEATURES             Location/Qualifiers
     source          1. .29276
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /strain="S288C (AB972)"
                     /db_xref="taxon:4932"
                     /map="VIII"
```

```
gene            275. .1273
                /gene="SNF6"
CDS             275. .1273
                /gene="SNF6"
                /note="YHL025w"
                /codon_start=1
                /product="Probable transcription factor"
                /protein_id="AAB65078.1"
                /db_xref="GI:2289895"
                /translation="MGVIKKKRSHHGKASRQQQYYSGVQVGGVGSMGAINNNIPSLTSF
                AEENNYQYGYSGSSAGMNGRSLTYAQQQLNKQRQDFERVRLRPEQLSNIIHDESDTIS
                FRSNLLKNFISSNDAFNMLSLTTVPCDRIEKSRLFSEKTIRYLMQKQHEMKTQAAELQ
                EKPLTPLKYTKLIAAAEDGSRSTKDMIDAVFEQDSHLRYQPDGVVVHRDDPALVGKLR
                GDLREAPADYWTHAYRDVLAQYHEAKERIRQKEVTAGEAQDEASLQQQQQQDLQQQQQ
                VVTTVASQSPHATATEKEPVPAVVDDPLENMFGDYSNEPFNTNFDDEFGDLDAVFF"
gene            2073. .4214
                /gene="YHL024w"
CDS             2073. .4214
                /gene="YHL024w"
                /note="Similiar to RNA binding proteins in N terminus with
                Asparagine rich region?"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAB65077.1"
                /db_xref="GI:2289894"
                /translation="MKTEISTADSLRDPPSNGLKADSELVIREDIDQFLPSEVSSLGS
                DHQNDGEDSDTDSDNFLQDPEDDVDEESTGRGTVTTTSTSTESRGRPSSCIFVASLAA
                ALSDDELCLSVTENFKKYGDLARVKVLRDNANRPYAFVQYNNDHDAKHALIRAQGTLL
                NGRRLRCEPAKVNRTLYLKNQQSIDFNEISQICEKFGGLEQIVPDRTDNQYTRRYTYP
                ISSANSWFVQPVYRDDAIRAYANLRTDPNWIIEWAQNINVPKNYNLLHKSKFKSSKYH
                QNNGIINNDGSNNNDNNNSNNNNREDSRRNGDVIEEECGHVHGSDSEEKLTSDGIYDD
                EDKDSEITIDKRSIFVGQLDKETTREELNRRFSTHGKIQDINLIFKPTNIFAFIKYET
                EEAAAAALESENHAIFLNKTMHVQYKEVGGRHNRKFSGKNGGSNFNHHQFFSTRSGKT
                FTGPELNLAPPPINMYRKMSGGSQQESETMMPYMPMGPMPMGPPPPNAASLSDFDMFP
                PSYSTFMKGMMPLRRKSMPNSWSSPSSKSVNSENESVNGGDENSELPSEIPESSGRYN
                AANSFTTYNNSSAGNSNNNNNNNSNSNKSQYKKRYARRSSYGYNEVPPKPYYFQPYY
                YHPMQYHMGPMGPLHPSQGSAGNHHPYMMVYPMSPPPPSGLDGSMIPPPINVSQSHAA
                NHGSTHVHANEFISNDTGDINEDNKAYSLDY"
gene            complement(4547. .7987)
                /gene="YHL023c"
CDS             complement(4547. .7987)
                /gene="YHL023c"
                /codon_start=1
                /evidence=not_experimental
                /product="unknown"
                /protein_id="AAB65076.1"
                /db_xref="GI:2289893"
                /translation="MDECLPNSCLLGVHLVISTHSGPQIVYHYPPSNTAFLTNNPTKH
                QHLYGNHANLNKNTSTNKEEKLFNSGSTKTASQIALNESAKSYNTAITPSMTNTNTNN
                VTLPPTRSHANTVGSQSSIPAATNGVGYRKTDIEDTSRTFQYQETESETSSSGLSDSE
                LSTDYLDISSDSFSISSSLSSSSLSSSPSSSSSSSPPQDGLSRTNSSFQSTDSMSPTS
                PQMIMENDSISVAESYLDSGTNNKSRAASKRSQNFFHKLSTKKSTDSKTHSPVRKLKS
                KPSQSTKKGNKLLKNTSNETDGNAFTGSCSISSKKSLSSTGEHNQELRNSSLNDTPGQ
                SPHHYHHRYHHYHKNAATSQRNSHTQYDVEEEDMEVSAMLQDGKISMNEIFFEEENFQ
                DINKILEFDNDFVAEFCSPEREMCNTRFEFTVDNFCFLGLPIHVDSQGRWRKSKHKNK
                TRSKRSSSTTTNISRKKSIASKISSLSENTLKKVNSGEADTVYDSNIGHEASTDTPNL
                RINTDVSGNEFEREKEDLGKNMNMFHVCFVMNPHLIEYNKRIDDMYQFVVTRLSLLLR
                YVQSKTSYISSECHIILKEKERVLKHSKTYQSIRGAGNKGKYLYQRILAKSSLARALT
                ECVDKIQRNEIACLEINDDKVISLQIPIQNEFEKMPNFKLQPVLRGSYLTSILNMKFL
                EKSSLRIESQNRQNDQAQFSDTNNNIYRFGNNINSTGHCGAANVDDGDDNESNYYCDD
                NDDLLNYALLLLDEPNNIISSLETFSYQDDIGTIILKHLVRNIQPNIPLRSYRYLITD
                LLDNPSSLDDLTTETNSLESSILRSCALHLMYWRHARIVIPLSSKYTYIVSPLAPIQG
                YTIDDYKSTSQNDGNVKKMDDRENNKSGSDRVPLIYQNSMLFRSKFPSLPSLPIFLSL
                LSTDKPQAYSNIIPSREHKPVYLNALAWLIQYGYVTQLLTFINIRVDKHIKMAVDEDL
                EKEGFRKTNTARRPSMDYKKTDKKLDDEDGQSRDANASEACSGKNEGMQSNDNNKDVD
                EKDNENDSRVDDRDDNEIAIADEEEILHFEYDDPEMQHDYTIILEPERATAIEKRWLY
                RCIYGQPSDIQILFNKLLKYFNGKVPMELVIIKEEISRHDLKKLLNALDKYLIEIHHW
                "
tRNA            8179. .8250
                /product="tRNA-His"
                /note="codon recognized: CAC"
gene            complement(8385. .9581)
                /gene="SPO11"
CDS             complement(8385. .9581)
                /gene="SPO11"
                /note="YHL022c"
                /codon_start=1
                /product="Sporulation protein"
                /protein_id="AAB65075.1"
                /db_xref="GI:2289892"
                /translation="MALEGLRKKYKTRQELVKALTPKRRSIHLNSNGHSNGTPCSNAD
                VLAHIKHFLSLAANSLEQHQQPISIVFQNKKKKGDTSSPDIHTTLDFPLNGPHLCTHQ
                FKLKRCAILLNLLKVVMEKLPLGKNTTVRDIFYSNVELFQRQANVVQWLDVIRFNFKL
                SPRKSLNIIPAQKGLVYSPFPIDIYDNILTCENEPKMQKQTIFPGKPCLIPFFQDDAV
                IKLGTTSMCNIVIVEKEAVFTKLVNNYHKLSTNTMLITGKGFPDFLTRLFLKKLEQYC
                SKLISDCSIFTDADPYGISIALNYTHSNERNAYICTMANYKGIRITQVLAQNNEVHNK
                SIQLLSLNQRDYSLAKNLIASLTANSWDIATSPLKNVIIECQREIFFQKKAEMNEIDA
                RIFEYK"
gene            complement(9885. .11282)
                /gene="YHL021c"
CDS             complement(9885. .11282)
                /gene="YHL021c"
                /codon_start=1
                /evidence=not_experimental
                /product="unknown"
                /protein_id="AAB65074.1"
                /db_xref="GI:2289891"
                /translation="MLRSNLCRGSRILARLTTTPRTYTSAATAAAANRGHIIKTYFNR
                DSTTITFSMEESSKPVSVCFNNVFLRDASHSAKLVTTGELYHNEKLTAPQDIQISEDG
                KSLVVKWKDGGHHQFPLQFFIDYKGSSFVSPATRKQESRYRPQLWNKRILKDNVKDLL
                SVSYNEFIDPKDDSKLFQTLVNLQKFGIAFISGTPSSSSEGLTIQKICERIGPIRSTV
                HGEGTFDVNASQATSVNAHYANKDLPLHTDLPFLENVPGFQILQSLPATEGEDPNTRP
                MNYFVDAFYATRNVRESDFEAYEALQIVPVNYIYENGDKRYYQSKPLIEHHDINEDNT
                LLGNYEALIKCINYSPPYQAPFTFGIYDKPSDLNNNLDLNLITTPAKLTERFLFKSFI
                RGLNLFESHINDFNNQFRLQLPENCCVIFNNRRILHANSLTSSNQQWLKGCYFDSDTF
                KSKLKFLEEKFPHDK"
gene            complement(11665. .12879)
                /gene="OPI1"
CDS             complement(11665. .12879)
                /gene="OPI1"
                /note="YHL020c"
                /codon_start=1
                /product="OPI1 (Negative regulator of phospholipid
                biosynthesis)"
                /protein_id="AAB65073.1"
                /db_xref="GI:2289890"
                /translation="MSENQRLGLSEEEVEAAEVLGVLKQSCRQKSQPSEDVSQADKMP
                ASESSTTPLNILDRVSNKIISNVVTFYDEINTNKRPLKSIGRLLDDDDDEHDDYDYND
                DEFFTNKRQKLSRAIAKGKDNLKEYKLNMSIESKKRLVTCLHLLKLANKQLSDKISCL
                QDLVEKEQVHPLHKQDGNARTTTGAGEDETSSDEDDDDEEFFDASEQVNASEQSIVVK
                MEVVGTVKKVYSLISKFTANSLPEPARSQVRESLLNLPTNWFDSVHSTSLPHHASFHY
                ANCEEQKVEQQQQQQQQQQQQLLQQQLLQQQQQKRNKDGDDSASPSSSVTANGKVLI
                LAKESLEMVRNVMGVVDSTLGKAEEWVKQKQEVKEMIRERFLQQQQQYRQQQQKDGNY
                VKPSQDNVDSKD"
gene            complement(13154. .14971)
                /gene="YHL019c"
CDS             complement(13154. .14971)
                /gene="YHL019c"
                /note="Similiar to medium chain of clathrin coat
                associated proteins"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAB65072.1"
                /db_xref="GI:2289889"
                /translation="MSSSLFILDENLEPLVSKNIRALPNLSSVLSSFKQCYHDGSPPI
                LSQNDWFFIHLKRDFLHFVSVIHTTDKPNIDLMTILAFLEQFYHLLQKYFEIEVLTKN
                VILDNILLVLELIDECIDFGIVQVTDPSIIKDYIRVKVNVPRVTVDNEEWSPGEESSS
                SSGSDSDSEYSNTNKRKDKKKKRKKKKGTKGKSVGKSKLKSIMVNNKENRGINVVETV
                KETLRNKNDTGKEAANDELPNDGNDLYINGDIAKTIIMPISWRTKGIHYAKNEFFLDV
                IERVQYLMDFEKGVIRKNLIHGEIVCRCYLSGMPKLKISINKILNRDPQFMSNSSFHQ
                CVSLDSINTIEKDEEKNSDDDAGLQAATDAREIEFIPPDGEFVLCQYELKRHVKDAPM
                VRLKDFEIKPKLKKFKIQIVTKIQTNFKPTNSTSKLNVRIPLTKVFQEYKIDLSKQIR
                FKANIGKVVFNLSDDFLLWEIQTMKGHREHSTNKSSQYNSDEDDPNTCASMVAEFPLF
                NQEEYDRLQEEMKTSMNPPPLRTGPRLEELYRQVHDQQTSHVTPRDKLVNIDFEIPYC
                TCSGLKVEYLKVEEPQLQYQSFPWVRYKTVSDEEYAYIV"
gene            15131. .15493
```

```
                     /gene="YHL018w"
     CDS             15131. .15493
                     /gene="YHL018w"
                     /note="Weakly similar to pterin-4-alpha-carbinolamine
                     dehydratase (human)."
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAB65071.1"
                     /db_xref="GI:2289888"
                     /translation="MHNKIVRIASSALTGGKLLEKLKPLTRWEVQWDPNKTKCLGITR
                     EVTFKDYETTWAFLTRVSMRSHLWGHHPLIHTSYTWVKLELHTHDIDPKDGAHSQLSD
                     IDVRMAKRIDSYIDEMTT"
     gene            15699. .17297

Alignment Scores:
Pred. No.:              6.04e+03       Length:         29276
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     15             Gaps:           0

US-09-858-332G-28 (1-6) x YSCH9433 (1-29276)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      11588 CATAATCATAATCATAAT 11571

RESULT 405
CQ590000/c
LOCUS       CQ590000               31034 bp    DNA     linear   PAT 02-FEB-2004
DEFINITION  Sequence 17758 from Patent WO0171042.
ACCESSION   CQ590000
VERSION     CQ590000.1  GI:41648646
KEYWORDS    .
SOURCE      Drosophila sp.
  ORGANISM  Drosophila sp.
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1
  AUTHORS   Venter,J.C., Adams,M., Li,P.W. and Myers,E.W.
  TITLE     Detection kits, such as nucleic acid arrays, for detecting the
            expression of 10,000 or more Drosophila genes and uses thereof
  JOURNAL   Patent: WO 0171042-A 17758 27-SEP-2001;
            PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1. .31034
                     /organism="Drosophila sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:7242"
ORIGIN

Alignment Scores:
Pred. No.:              6.39e+03       Length:         31034
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     6              Gaps:           0

US-09-858-332G-28 (1-6) x CQ590000 (1-31034)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      19202 CATAATCATAATCATAAT 19185

RESULT 406
AB013387/c
LOCUS       AB013387               31309 bp    DNA     linear   PLN 14-FEB-2004
DEFINITION  Arabidopsis thaliana genomic DNA, chromosome 5, TAC clone:K18G13.
ACCESSION   AB013387 BA000015
VERSION     AB013387.1  GI:3128134
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Kotani,H., Nakamura,Y., Sato,S., Asamizu,E., Kaneko,T., Miyajima,N.
            and Tabata,S.
  TITLE     Structural analysis of Arabidopsis thaliana chromosome 5. VI.
            Sequence features of the regions of 1,367,185 bp covered by 19
            physically assigned P1 and TAC clones
  JOURNAL   DNA Res. 5 (3), 203-216 (1998)
   PUBMED   9734815
REFERENCE   2  (bases 1 to 31309)
  AUTHORS   Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-1998) Yasukazu Nakamura, Kazusa DNA Research
            Institute, Department of Plant Gene Research; 1532-3, Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
            Tel:81-438-52-3935, Fax:81-438-52-3934)
COMMENT     Address for correspondence: kaos@kazusa.or.jp
            For the latest information on annotation of this clone, please see
            http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=K18G13
            Genes with similarity to proteins in the databases are described in
            'product' or 'note' qualifiers.  Genes that have no significant
            protein similarity are described as 'unknown protein'.
            The software programs used to predict genes include: Grail
            (Informatics Group, Oak Ridge National Laboratory,
            http://compbio.ornl.gov/Grail-1.3/),
            GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
            NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
            Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
            SplicePredictor (Volker Brendel, Stanford University,
            http://gremlin1.zool.iastate.edu/cgi-bin/sp.cgi).
            Genes encoding tRNAs are predicted by tRNAscan-SE
            (Sean Eddy, Washington University School of Medicine, St. Louis,
            http://genome.wustl.edu/eddy/tRNAscan-SE/).
            This sequence may not be the entire insert of this clone. It may be
            shorter because we remove overlaps between neighboring submissions.
            The 5' clone is MJP23 and the 3' clone is MDK4.
FEATURES             Location/Qualifiers
     source          1. .31309
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /chromosome="5"
                     /clone="K18G13"
                     /clone_lib="Mitsui TAC"
                     /ecotype="Columbia"
     CDS             complement(join(10536. .10643,10897. .11121,12045. .12308))
                     /note="unnamed protein product; gene_id:K18G13.2
                     pir||T05871
                     similar to unknown protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAB11577.1"
                     /db_xref="GI:10178164"
                     /translation="MFGLPLFPSDLYELGKMGTKARTTKPSLEEEGSPLRGILCVKTR
                     RDMKRFEKTEDCFILDFNPFDSFDVKRLSFSLDKDLEIIHESGQTFVAKKMEKKEEDD
                     TPLVGIFCLKTRQDMKRFEETEDCFILDFDPFDSFDVKKLTFTDDVREGDKDLSIIHE
                     TGQCYCYVCDKPAPCAEWTISTSHCNAYDSALWNLCRT"
     CDS             join(16146. .16558,17365. .17675,17756. .17820,17936. .18238)
                     /note="gene_id:K18G13.3"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="O-methyltransferase"
                     /protein_id="BAB11578.1"
                     /db_xref="GI:10178165"
                     /translation="MGSTAETQLTPVQVTDDEAALFAMQLASASVLPMALKSALELDL
                     LEIMAKNGSPMSPTEIASKLPTKNPEAPVMLDRILRLLTSYSVLTCSNRKLSGDGVER
                     IYGLGPVCKYLTKNEDGVSIAALCLMNQDKVLMESWYHLKDAILDGGIPFNKAYGMSA
                     FEYHGTDPRFNKVFNNGMSNHSTITMKKILETYKGFEGLTSLVDVGGGIGATLKMIVS
```

```
                KYPNLKGINFDLPHVIEDAPSHPGIEHVGGDMFVSVPKGDAIFMKWICHDWSDEHCVK
                FLKNCYESLPEDGKVILAECILPETPDSSLSTKQVVHVDCIMLAHNPGGKERTEKEFE
                ALAKASGFKGIKVVCDAFGVNLIELLKKL"
CDS             complement(join(19127. .19620,19706. .19787,20058. .20141,
                20245. .20421,20507. .20653,20892. .21110))
                /note="gene_id:K18G13.4"
                /codon_start=1
                /evidence=not_experimental
                /product="membrane related protein-like"
                /protein_id="BAB11579.1"
                /db_xref="GI:10178166"
                /translation="MALLESIILGLITGWLAFVIGLVIGWAWKPRWVSSSNNRQVKLQ
                CSAPRSFDLSLPSSSPSSVVTSPLKGFGSERLPNTVTELDLRHLVQLVERKDGGQAWI
                QMMDRFTPGMRYQAWLREPKNGPTEYRSRTVFEDATPVMLRDFFWDDEFRPTWDTMLS
                NSTTVEECPSTGTMIVRWIRKFPFFCSDREYVIGRRIWNCGNSYYCVTKGVSVPSIPP
                NNKQKRVDLFYSSWCIRPVESRRDDGVTSACEVLLFHHEDMGIPREIAKLGVKRGMWG
                AVKKMEPGLRAYQTHRLSDGGTKLSRSAFMAQINTKITQDHLISLSNGASPVTETPVT
                LNQGNRAANLKKLLIIGGAVAVVCSLSGGAFVPPAFLLGFGKRFVNGGRKRQLEGTTR
                SESQTTSS"
CDS             join(22615. .22849,22987. .24254)
                /note="unnamed protein product; gene_id:K18G13.5
                pir||T05643
                similar to unknown protein"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="BAB11580.1"
                /db_xref="GI:10178167"
                /translation="MVILSLVSCSFSVFSPPISLRLHLPPVTSLCSHGTFPASSTFRS
                QLQPLLISCLNHREPALTFRCSCLSSPIESGSQIESLFSLFRDIGFIEEETEMILAKN
                PDIKSTSLDKIGARVASLQSLKINGFPLQGLIAKCPNLLTSEEFDLVISFLVDELEGR
                LDPELVERLLSVVDTSILLSFNQKVRLLLLHGIPKEKISHVLNKVYLNKLLYQKSVED
                IERLISFLEPFGGIGIIARRPVILNSDLDSQLIPRVDFIRNLSGEDDFATGTVLRRLP
                AILSYSVEHMNGQVEFLKSFAGLTSEQVFKIVHVFPNVISTSKERKLRPRIEFLKECG
                FDSPGMFKFLSKAPLILALSENNLSHKLGFLVKIGYKHRTKELAFAMGAVTRTSSDNM
                QRVIGLYLSYGLSFEDILAMSTKHPQVLQYNYTSLEEKLEYLIEYMGREVEELLAFPA
                FLGYKLDSRIKHRYEEKLKSRGENMSLNKLLTVSAERFSKAADNIEMICL"
snRNA           24663. .24881
                /product="U3 snRNA"
                /note="gene_id:K18G13.6"
                /evidence=not_experimental
CDS             complement(join(25254. .25409,25514. .25872,25970. .26507,
                26590. .26700,26791. .26844))
                /note="gb|AAC49043.1
                gene_id:K18G13.7"
                /codon_start=1
                /evidence=not_experimental
                /product="NADPH:protochlorophyllide oxidoreductase A"
                /protein_id="BAB11581.1"
                /db_xref="GI:10178168"
                /translation="MALQAASLVSSAFSVRKDGKLNASASSSFKESSLFGVSLSEQSK
                ADFVSSSLRCKREQSLRNNKAIIRAQAIATSTPSVTKSSLDRKKTLRKGNVVVTGASS
                GLGLATAKALAETGKWHVIMACRDFLKAERAAQSAGMPKDSYTVMHLDLASLDSVRQF
                VDNFRRAEMPLDVLVCNAAVYQPTANQPTFTAEGFELSVGINHLGHFLLSRLLIDDLK
                NSDYPSKRLIIVGSITGNTNTLAGNVPPKANLGDLRGLAGGLNGLNSSAMIDGGDFVG
                AKAYKDSKVCNMLTMQEFHRRFHEDTGITFASLYPGCIATTGLFREHIPLFRTLFPPF
                QKYITKGYVSESEAGKRLAQVVADPSLTKSGVYWSWNKTSASFENQLSQEASDVEKAR
                RVWEVSEKLVGLA"
CDS             complement(join(27636. .28195,28440. .28533,28644. .28726,
                28814. .28865,28957. .29114,29380. .29470,29708. .31147))
                /note="unnamed protein product; contains similarity to
                WD-containing protein
                gene_id:K18G13.8"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="BAB11582.1"
                /db_xref="GI:10178169"
                /translation="MMKMMMMNRRRSNVVRFEAKKEDKVEEEDNNDDCFYESLDRVLS
                SCSCSTSNSDYDSDPNVSDSIHDPTPFPVPMFTSGFELWKSEPESVSERRIRLLRGLG
                LSNEPDLPPASQLRSRSRRNSSHFSRSVPSDVLTSNHHGKCVLRSDVSGTDNVDREYT
                GSRMRNDGVDFISKEPMERSLRDIVEKENGRDVVLEEQMCTIRNLDTGREFVVNEVRE
                DGMLERLKEVGTDRQLTLEEFEICVGTSPIVLELMRRQNVEDVCKDSVDLDTNVSGSR
                VTKHRRRGSWLKSIKNVASSVTGYKERRSTDDRDSPSERGGQRFSSATDDSRDMSFHD
                PERVKVRQYGKSCKELTALFKSQEIQAHKGSIWSIKFSLDGRYLASAGEDCVIQIWKV
                VESERKGELLSMDKQEDGSINLFLLANGSPEPVSMSPKRRGRTSFSRKSVSLDNVLVP
                EAVFGLSEKPVCSFVGHLDDVLDLSWSKSQHLLSSSMDKTVRLWDLSSKTCLKVFSHS
                DYVTCIQFNPVDDNYFISGSLDAKVRIWSIPDHQVVDWNDLHEMVTAACYTPDGQGAL
                VGSYKGTCCLYNTHDNKLQQRREINLKNRKKKTHHKKITGFQFVAGSSSEVLVTSADS
                RTRVVDGVDLVHKFKGFRNTNSQISASLTSNGKFLVSASEDSNVYVWNYDSDSRAGRS
                KRVTVTNSYENFYCRDVSVAAPWPGKISNNHTNNSPEQSSFTANNNPPTPVNDPTNNK
                TVTNGIISSATNRYFFDRVSATWPEEKLLLAAKNRARTSPRVSVDMSNIQVNTKPSAS
                AWSMVIVTGSLRGEIRIFQNFGFPVRL"
ORIGIN

Alignment Scores:
Pred. No.:                6.45e+03        Length:          31309
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       15              Gaps:            0

US-09-858-332G-28 (1-6) x AB013387 (1-31309)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      14048 CACAACCACAACCACAAC 14031

RESULT 407
AF016437/c
LOCUS       AF016437               31361 bp    DNA     linear   INV 21-SEP-2004
DEFINITION  Caenorhabditis elegans cosmid F13H6, complete sequence.
ACCESSION   AF016437
VERSION     AF016437.1  GI:2315331
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 31361)
  AUTHORS   .
  CONSRTM   WormBase Consortium
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
  JOURNAL   Science 282 (5396), 2012-2018 (1998)
   PUBMED   9851916
REFERENCE   2  (bases 1 to 31361)
  AUTHORS   Jones,K. and Wohldmann,P.
  TITLE     The sequence of C. elegans cosmid F13H6
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 31361)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUL-1997) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 31361)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   5  (bases 1 to 31361)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-2002) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   6  (bases 1 to 31361)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2002) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   7  (bases 1 to 31361)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-OCT-2003) Department of Genetics, Washington
```

```
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   8  (bases 1 to 31361)
  AUTHORS   .
  CONSRTM   WormBase Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (21-SEP-2004) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
COMMENT     Submitted by:
                  Genome Sequencing Center
                  Department of Genetics, Washington University
                  St. Louis , MO  63110, USA, and
                  Sanger Centre, Hinxton Hall
                  Cambridge CB10 IRQ, England
                  email: submissions@watson.wustl.edu and jes@sanger.ac.uk

            NOTICE:  This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate chemistry
            or covered by high quality data (i.e., phred quality >= 30); an
            attempt was made to resolve all sequencing problems, such as
            compressions and repeats; all regions were covered by sequence from
            more than one m13 subclone.

            For a graphical representation of this clone sequence and its
            analysis see:
            http://www.wormbase.org/db/seq/sequence?name=F13H6;class=Sequence

                        NEIGHBORING CLONE INFORMATION

            The 5' clone is T28C12, 200 bp overlap; the 3' clone is F57F4, 200
            bp overlap. Actual start of this clone is at base position 1 of
            F13H6; actual end is at 9073 of F57F4.

            NOTES:

            Coding seqences below are the result of integration and manual
            review of the following data : computer analysis using the program
            Genefinder (P. Green and L. Hillier, personal communication), the
            large scale EST projects of Yuji Kohara
            (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C.
            elegans ORFeome cloning project (http://worfdb.dfci.harvard.edu/),
            similarity to other proteins from BlastX analyses
            (http://blast.wustl.edu/), sequence conservation with C. briggsae
            using Jim Kent's WABA alignment program (Genome Research
            10:1115-1125, 2000), individual C. elegans GenBank submissions,
            and personal communications with C. elegans researchers. tRNAs
            are predicted using the program tRNAscan-SE (Lowe, T.M. and
            Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).
FEATURES            Location/Qualifiers
     source         1. .31361
                    /organism="Caenorhabditis elegans"
                    /mol_type="genomic DNA"
                    /strain="Bristol N2"
                    /db_xref="taxon:6239"
                    /chromosome="V"
                    /clone="F13H6"
     gene           11642. .15769
                    /gene="F13H6.1"
                    /locus_tag="F13H6.1"
     CDS            join(11642. .11656,11708. .11813,11861. .11905,12424. .12638,
                    12769. .12870,13178. .13259,14111. .14834,14886. .15361,
                    15412. .15683,15734. .15769)
                    /gene="F13H6.1"
                    /locus_tag="F13H6.1"
                    /standard_name="F13H6.1"
                    /note="contains similarity to Pfam domain PF00096 (Zinc
                    finger, C2H2 type); coded for by the following C. elegans
                    cDNAs: yk5e12.5, yk6g5.5, yk6h10.5, yk13e9.5, yk13f5.5,
                    yk35h3.5, yk287c5.5, yk316d11.5, yk397g6.5, yk796g03.5,
                    yk1193d11.5, yk1197b08.5, yk1201a08.5, yk1202a07.5,
                    CEESA85F, CEESW01F, yk5e12.3, yk6g5.3, yk6h10.3, yk13e9.3,
                    yk13f5.3, yk287c5.3, yk397g6.3, yk756a01.3, yk784b06.3,
                    yk796g03.3, yk1135e02.3, yk1193d11.3, yk1197b08.3,
                    yk1202a07.3, yk1208c12.3"
                    /codon_start=1
                    /product="Hypothetical protein F13H6.1"
                    /protein_id="AAB65884.2"
                    /db_xref="GI:20901943"
                    /db_xref="WormBase:F13H6.1"
                    /translation="MLHLPTIETSSAGRRRRKANSEVDSGAEEPTATSSSDDKSDYRL
                    DLKKRLTTFLTSNSQMASSDSGVSGVESTVEHDLSELKASDSPMSHDGADSPRPELDR
                    TDIVVCGECHTSFSLSCFSNFIEHKVSNCGGKLTPSDDCEVTPRNFDRSRRRTFNPAA
                    LIRCGRSTSANALLPNYNMDVTTDTNDLGSSSKSLTCCNCKERFGDVWTLLKHSYDIH
                    GLRVCQEAIPESDVSVTSTSSPTSNESMHMSSRYNHNDTSSKIMHQTIKSESTMMNSF
                    CSERLKEIVTRAGEQDQKPLMALRRLMPSEETEEEQVSAFTSTSLLNSRASQPQLPTV
                    SSSAGFQTSGSMPNVWMQPSMLAAMQEYYASIQQMPYPMTNSTAVALLNISNNLQQQQ
                    QQQQQLQQQQQNLIAAQPQVATPQPPAPTPLFQSLNRLSTADESSYTPRPASVAIRRR
                    ASPEEELTPPRKSMKIEEEDQLIVVDDGELAEPAARKPNNIKKERCNFCNKVFTNRSN
                    LIVHLRSHTGEKPYKCQLCPYACAQSSKLTRHMRTHGQQGKETYHCYICRMPFSVHST
                    LEKHMRKCVVSTQGRREGGSPPERSIRPTPSALAEATSLLALSAPSISVPPQPQASTV
                    GQNQSQSHQIVLNWLHALNASNSTPTATSKEDMIEADEDMEETEASDLQSHQPLTSES
                    NQIATVAN"
     gene           complement(17546. .19701)
                    /gene="F13H6.3"
                    /locus_tag="F13H6.3"
     CDS            complement(join(17546. .17753,17797. .17926,17973. .18626,
                    18703. .18947,19081. .19169,19220. .19306,19402. .19560,
                    19609. .19701))
                    /gene="F13H6.3"
                    /locus_tag="F13H6.3"
                    /standard_name="F13H6.3"
                    /note="contains similarity to Pfam domain PF00135
                    (Carboxylesterase); coded for by the following C. elegans
                    cDNAs: OSTF021H2_1, OSTR021H2_1, yk552g2.3, yk552g2.5,
                    yk709f7.3, yk709f7.5, yk815d03.3, yk815d03.5, cm06b1.5,
                    yk1496h03.3, yk1496h03.5"
                    /codon_start=1
                    /product="Hypothetical protein F13H6.3"
                    /protein_id="AAB65887.1"
                    /db_xref="GI:2315336"
                    /db_xref="WormBase:F13H6.3"
                    /translation="MGGFLSHLTPEQNVEALKASCGPVRGNIYKHDDVIVDGYLGIPY
                    AKPPVGELRFKKPVTVDVWTEIKDCYKYGPACVQTGGFEQIAGPRTPTPEEAGCLTLN
                    VFTPRNASSEFKNGRPVMVYIHGGGYELCASSDFCAYSLSGTLPLKDVVVVSINYRLG
                    VFGFLTTGDNVCPGNFGLWDQTLALKWVQKHISSFGGDPNCVTVFGQSAGGASTDLLS
                    LSPHSRDLFQRFIPISGTAHCDFAIRASENQAKIFREFAEFHGFSGRDSSALFKWYQE
                    QSPETLSNVKGYKKSISGFLTFIPNLDGDFFPKPLDELRKEAPKKQMMTGVTEYEGLM
                    LASMNPAFSPADVGLTLMPQGIYGKDVVSNPDEIQKIFYEKYVEGVDKSDELAMRKKL
                    CEALGDEFFNVGVIQAAKNAAKHGNEVYFYTFEYVNPDSFGMWDGMMPFKAAVHCTEL
                    RYLLGEGVYSKFEPTEEDRKVMETTTTLFSNFAKYGNPNGKGATAEIWEKYSLNRPER
                    HYRISYPKCEMRDVYHEGRIQFLEKIDGDSDKYQELVYGKKKSAKI"
     misc_feature   complement(19704. .19705)
                    /note="SL1 trans-splice acceptor; see cm06b1.5"
     gene           complement(20466. .22045)
                    /gene="F13H6.4"
                    /locus_tag="F13H6.4"
     CDS            complement(join(20466. .20673,20792. .20921,20968. .21352,
                    21441. .21619,21687. .21931,21975. .22045))
                    /gene="F13H6.4"
                    /locus_tag="F13H6.4"
                    /standard_name="F13H6.4"
                    /note="contains similarity to Pfam domain PF00135
                    (Carboxylesterase); coded for by the following C. elegans
                    cDNAs: OSTF186A8_1, OSTR186A8_1"
                    /codon_start=1
                    /product="Hypothetical protein F13H6.4"
                    /protein_id="AAB65885.2"
                    /db_xref="GI:37515241"
                    /db_xref="WormBase:F13H6.4"
                    /translation="MVFIHGGGFEFGGSADYCDYSLSGTLPLKDVVVVSINYRLGVFG
```

```
                     FLTTGDSVCNGNFGLWDQTLALKWVQKHISSFGGDPNCVTVFGQSAGGASTDLLSLSP
                     HSRDLFQRFIPISGSAYCEFALRTSKSQAKIFREFAEFKGFTGGDSTTLLEWYKNQSS
                     ETLSDLRKEAPKKQMMTGVDEYEGVIASMNPEFSPADAGLALFPKGVYGNDTAENPEE
                     MHKLFYEKYVEGVDKSDDSAMKKRLCEAFGDLGFNLGVFQSAKSSAKYGNDVFLYSFE
                     YYHSDGFGMWKDLLPFEASMHGTELRYLLGEGFYSKFDATKEELEVLEKTTTLFSNFA
                     KYGNPNGTGVTDEIWEKFNLNNPERHFRISYPKCEMRDIFSGGRMPFLEKLEEKSLKT
                     QELVYGTKKFTNN"

Alignment Scores:
Pred. No.:                6.46e+03        Length:          31361
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       2               Gaps:            0

US-09-858-332G-28 (1-6) x AF016437 (1-31361)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      23674 CATAATCATAATCATAAT 23657

RESULT 408
AY292861
LOCUS       AY292861               31901 bp    DNA     linear   PLN 29-MAR-2004
DEFINITION  Boechera drummondii zing finger transcription factor PEI1 (PEI1),
            pollen specific glycine-rich protein GRP17 (GRP17), pollen specific
            glycine-rich protein GRP16 (GRP16), pollen specific glycine-rich
            protein GRP19 (GRP19), pollen specific glycine-rich protein GRP20
            (GRP20), transcription-factor-like protein (Bd5g07580), and WD
            repeat-like protein (Bd5g07590) genes, complete cds; and oleosin
            (Bd5g07600) and F box protein family (Bd5g07610) genes, partial
            cds.
ACCESSION   AY292861
VERSION     AY292861.1  GI:34013884
KEYWORDS    .
SOURCE      Boechera drummondii
  ORGANISM  Boechera drummondii
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Boechera.
REFERENCE   1  (bases 1 to 31901)
  AUTHORS   Schein,M., Yang,Z., Mitchell-Olds,T. and Schmid,K.J.
  TITLE     Rapid Evolution of a Pollen-Specific Oleosin-Like Gene Family from
            Arabidopsis thaliana and Closely Related Species
  JOURNAL   Mol. Biol. Evol. 21 (4), 659-669 (2004)
   PUBMED   14739246
REFERENCE   2  (bases 1 to 31901)
  AUTHORS   Schmid,K.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAY-2003) Department of Genetics and Evolution,
            Max-Planck-Institut of Chemical Ecology, Winzerlaer Str. 10, Jena
            07745, Germany
FEATURES             Location/Qualifiers
     source          1. .31901
                     /organism="Boechera drummondii"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:72658"
     gene            <1. .>741
                     /gene="PEI1"
                     /locus_tag="Bd5g07500"
     mRNA            <1. .>741
                     /gene="PEI1"
                     /locus_tag="Bd5g07500"
                     /product="zing finger transcription factor PEI1"
     CDS             1. .741
                     /gene="PEI1"
                     /locus_tag="Bd5g07500"
                     /codon_start=1
                     /product="zing finger transcription factor PEI1"
                     /protein_id="AAQ56110.1"
                     /db_xref="GI:34013885"
                     /translation="MLKSASPLAFYDIGDQQYSTLGYIVSKPGIAEAYDLDHPIPTIN
                     NAIYGSDEFRMVAYKIKRCPRTRSHDWTECPYAHRGEKATRRDPRRYTYCAVACPAFR
                     NGACHRGDSCEFAHGVFEYWLHPARYRTRACNAGNMCQRKVCFFAHAPEQLRQSEGKH
                     RCRYAYRPVRARGGNGDGVAMRLDGEDYDTSRSPEKSGKGDLDSNEEKVLLKCWTSMT
                     IVDDHYQPSDLDLDLDLSHIDWISELVD"
     gene            complement(<6632. .>8202)
                     /gene="GRP17"
                     /locus_tag="Bd5g07530"
     mRNA            complement(join(<6632. .7247,7796. .>8202))
                     /gene="GRP17"
                     /locus_tag="Bd5g07530"
                     /product="pollen specific glycine-rich protein GRP17"
     CDS             complement(join(6632. .7247,7796. .8202))
                     /gene="GRP17"
                     /locus_tag="Bd5g07530"
                     /codon_start=1
                     /product="pollen specific glycine-rich protein GRP17"
                     /protein_id="AAQ56111.1"
                     /db_xref="GI:34013886"
                     /translation="MSEELTQNPSLTQSLSQREGRKIHISQREGRKFPFQYLSEREER
                     KLSFLNMFSFLMPLLEVIKIIIASVASVVFLGFAGVTLAGSAVALAVSTPLFIIFSPV
                     LVPATIATVVLTTGFTAGGSFGATALGLIVWLIKRRMGVPPKDNPPPKELPPNAGGGG
                     ASAQTLLEKPKSKGGFKLPDWFKKIPGLGGIFGGGKGRFSPGGEGLFGGRGGKLKEKA
                     RKCKGRKGKSKSKGGMSGGGGGMSEGEGGMSSGGEGMSGSGGAKSKGEGGMSGSEGGM
                     SGGGMSGGGMSGSEGGVSGSEGCKHGGGGGKHGGGGGKHEGKHGCKHGGGGGKHGGRG
                     SDVEEE"
     gene            complement(10554. .11365)
                     /gene="GRP16"
                     /locus_tag="Bd5g07540"
     mRNA            complement(join(<10554. .10860,11109. .>11365))
                     /gene="GRP16"
                     /locus_tag="Bd5g07540"
                     /product="pollen specific glycine-rich protein GRP16"
     CDS             complement(join(10554. .10860,11109. .11365))
                     /gene="GRP16"
                     /locus_tag="Bd5g07540"
                     /codon_start=1
                     /product="pollen specific glycine-rich protein GRP16"
                     /protein_id="AAQ56113.1"
                     /db_xref="GI:34013888"
                     /translation="MLSFLIPVVQLFQVVIAAVASVVFLVFAGITLGASVVGLTIATP
                     LFVIFSPILVPATITTTLLVVGVVGTAALGVAAIALILWLLKHRIGVKPTNNPAPEEA
                     PTKADKPAGDKPSGAPGDNAGGAPGDKAGGASGDKPAGASGDKPGEAPEDKPGEAPGD
                     KPGDKPGGASGDKPGGAPGDKQPACKG"
     gene            complement(<13308. .>13935)
                     /gene="GRP19"
                     /locus_tag="Bdg07550"
     mRNA            complement(join(<13308. .13380,13700. .>13935))
                     /gene="GRP19"
                     /locus_tag="Bdg07550"
                     /product="pollen specific glycine-rich protein GRP19"
     CDS             complement(join(13308. .13380,13700. .13935))
                     /gene="GRP19"
                     /locus_tag="Bdg07550"
                     /codon_start=1
                     /product="pollen specific glycine-rich protein GRP19"
                     /protein_id="AAQ56112.1"
                     /db_xref="GI:34013887"
                     /translation="MFEIVQAVFAAGGALALLTFAGITLGGSVVAFVISTPLFVIFSP
                     VLVPATIATTLLATGFTASGSLGSTAISILMWLYKKRTGKDPPKIPGLTPDSAPGKYD
                     "
     gene            complement(<15183. .>16380)
                     /gene="GRP20"
                     /locus_tag="Bd5g07560"
     mRNA            complement(join(<15183. .15285,16157. .>16380))
                     /gene="GRP20"
                     /locus_tag="Bd5g07560"
                     /product="pollen specific glycine-rich protein GRP20"
     CDS             complement(join(15183. .15285,16157. .16380))
                     /gene="GRP20"
                     /locus_tag="Bd5g07560"
                     /codon_start=1
                     /product="pollen specific glycine-rich protein GRP20"
                     /protein_id="AAQ56114.1"
```

```
                     /db_xref="GI:34013889"
                     /translation="MTAFLATTAAVFLLLLAGLSLTGTALALVASMPLFLLFSPVLVP
                     AGITTTVLASGLMAGGTSGVTALNIIMWLYKQFTGKDPPKIPGLTPAGGAASAKKPAA
                     KPAASG"
     gene            <21711. .>22337
                     /locus_tag="Bd5g07580"
     mRNA            <21711. .>22337
                     /locus_tag="Bd5g07580"
                     /product="transcription-factor-like protein"
     CDS             21711. .22337
                     /locus_tag="Bd5g07580"
                     /codon_start=1
                     /product="transcription-factor-like protein"
                     /protein_id="AAQ56115.1"
                     /db_xref="GI:34013890"
                     /translation="MATIEESSDLEAIQKHLFEDLMVSDGFMGDFDFDASFVSELWCI
                     EPQVPKQEPDSPVFDPDSFVNEFLQVEAKSSPELNSSSSTYETDLSVKKAERFEEEEE
                     EEEEEARHYRGVRRRPWGKFAAEIRDPAKKGSRIWLGTFESDVDAARAYDCAAFKLRG
                     RKAVLNFPLEAGKYEAPANSGRKRKRSDVKEELQRTQSNSSSSSCDAF"
     gene            complement(<23559. .>25625)
                     /locus_tag="Bd5g07590"
     mRNA            complement(join(<23559. .23926,24039. .24146,24250. .24322,
                     24428. .24604,24759. .25247,25509. .>25625))
                     /locus_tag="Bd5g07590"
                     /product="WD repeat-like protein"
     CDS             complement(join(23559. .23926,24039. .24146,24250. .24322,
                     24428. .24604,24759. .25247,25509. .25625))
                     /locus_tag="Bd5g07590"
                     /codon_start=1
                     /product="WD repeat-like protein"
                     /protein_id="AAQ56116.1"
                     /db_xref="GI:34013891"
                     /translation="MEFTEAYKQTGPCCFSPNSRYVAVANDYRLVIRDTFSFQVVQLF
                     SCLDKISYIEWALDSEYILCGLYKKPMIQAWSLTQPEWTCKIDEGPAGISYARWSPDS
                     RHILTTSEFQLRLTVWSLLNTACVHVQWPKHGSKGVAFNQDGKFAAICTRRDCKDYVN
                     LLSCQSWEIMGSFAVDTLDLADLEWSPDDSSIVVWDSPLEYKAYECGLGVKTVSWSPC
                     GQFLAIGSYDQMLRVLNHLTWKTFAEFLHLSTVRAPCSAAIFKEIDEPLQLDMSELSL
                     DENFMPSNYDASEGYISVRYEVMELPVAFPFQKPPADKPNPKQGVGLLAWSKDSQYIC
                     TRNDNMPTALWIWDMCHLEVAAILVQKEPIRAAVWDPTCTRLLLCTGSSHLYMWTPSG
                     AFCVCNPLPGFSISDLKWNIDGSCLLLKDKDAFCCATVPSLPESSDYSSDD"
     gene            complement(<29896. .>30145)
                     /locus_tag="Bd5g07600"
     mRNA            complement(<29896. .>30145)
                     /locus_tag="Bd5g07600"
                     /product="oleosin"
     CDS             complement(<29896. .30145)
                     /locus_tag="Bd5g07600"
                     /note="start codon not determined"
                     /codon_start=1
                     /product="oleosin"
                     /protein_id="AAQ56117.1"
                     /db_xref="GI:34013892"
                     /translation="MFSFLILFLEVFQVVISVVASIVFLVLAGLTLAGSAVALSITTP
                     LFIIFSPILVPATIATAVLTTGFTAGSTLGATTIILLSS"
     gene            <31650. .>31901
                     /locus_tag="Bd5g07610"
     mRNA            <31650. .>31901
                     /locus_tag="Bd5g07610"
                     /product="F box protein family"
     CDS             <31650. .31901
                     /locus_tag="Bd5g07610"
                     /note="start codon not determined"
                     /codon_start=1
                     /product="F box protein family"
                     /protein_id="AAQ56118.1"
                     /db_xref="GI:34013893"
                     /translation="MIRAENFTNRRIYAFSIIGFVKEETDAESYIVLHIPAKALKYNF
                     INRTFKKLCDFKPLVNNAPEDEFYRFQRTFQFIESLANV"
ORIGIN

Alignment Scores:
Pred. No.:        6.57e+03      Length:       31901
Score:            42.00         Matches:      6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     15            Gaps:          0

US-09-858-332G-28 (1-6) x AY292861 (1-31901)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       3536 CACAATCACAACCACAAC 3553

RESULT 409
AF309805
LOCUS       AF309805               32083 bp    DNA     linear   PLN 14-MAR-2001
DEFINITION  Pneumocystis carinii f. sp. carinii glutathione synthetase Gsh1
            (gsh1) gene, partial cds; putative a-factor pheromone receptor Ste3
            (ste3) gene, complete cds, alternatively spliced; and Ydr124wp-like
            protein, helix-loop-helix DNA binding protein, Yer049wp-like
            protein, protein kinase Ste20 (ste20), putative sucrose carrier
            Sca1 (sca1), putative Na+/H+ exchanger Nhx1 (nhx1), putative 26S
            protease regulatory subunit 4 (mts2), hypothetical protein, glycyl
            tRNA synthetase, and coiled-coil protein genes, complete cds.
ACCESSION   AF309805
VERSION     AF309805.1  GI:11596246
KEYWORDS    .
SOURCE      Pneumocystis carinii
  ORGANISM  Pneumocystis carinii
            Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;
            Pneumocystidaceae; Pneumocystis.
REFERENCE   1  (bases 1 to 32083)
  AUTHORS   Smulian,A.G., Sesterhenn,T., Tanaka,R. and Cushion,M.T.
  TITLE     The ste3 pheromone receptor gene of Pneumocystis carinii is
            surrounded by a cluster of signal transduction genes
  JOURNAL   Genetics 157 (3), 991-1002 (2001)
   PUBMED   11238389
REFERENCE   2  (bases 1 to 32083)
  AUTHORS   Smulian,A.G., Sesterhenn,T.M., Tanaka,R. and Cushion,M.T.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-OCT-2000) Infectious Disease Division, Department of
            Internal Medicine, University of Cincinnati College of Medicine,
            231 Bethesda Avenue, Cincinnati, OH 45267-0560, USA
FEATURES             Location/Qualifiers
     source          1. .32083
                     /organism="Pneumocystis carinii"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4754"
                     /clone="cosmid W15A6"
                     /note="forma specialis: carinii"
     gene            <1. .900
                     /gene="gsh1"
     mRNA            join(<46. .339,379. .701,753. .900)
                     /gene="gsh1"
                     /product="glutathione synthetase Gsh1"
                     /note="5' boundary not determined"
     CDS             join(<46. .339,379. .701,753. .844)
                     /gene="gsh1"
                     /note="5' boundary not determined"
                     /codon_start=2
                     /evidence=not_experimental
                     /product="glutathione synthetase Gsh1"
                     /protein_id="AAG38537.1"
                     /db_xref="GI:11596248"
                     /translation="SHGIQAYRVSLHSILDSIELRQNDKALLFKSPNSKIEEISVVYF
                     RTGYSPSHYPTSIQWSARLLIEESRAIKCPTVITQLAGCKKVQQVLCENNIINKFLPE
                     NISKSLKETFVSIYPLNKSLSGQEAYHLLMNNPQNYVLKPQREGGGNNIYGLNILSFL
                     ETISESQREQYILMEMIHPFLQTNTIVQNNELYHTNGGNELGIYGGYLLRTKDKDKNE
                     GGVMIGCSSIDSCVLIP"
     gene            1058. .2865
                     /gene="ste3"
     mRNA            join(1058. .1887,1934. .2411,2463. .2865)
                     /gene="ste3"
                     /product="putative a-factor pheromone receptor Ste3a"
     mRNA            join(1058. .1887,1934. .>2415)
```

/gene="ste3"
/product="putative a-factor pheromone receptor Ste3b"
CDS join(1359. .1887,1934. .2411,2463. .2775)
/gene="ste3"
/note="alternatively spliced"
/codon_start=1
/evidence=experimental
/product="putative a-factor pheromone receptor Ste3a"
/protein_id="AAG38536.1"
/db_xref="GI:11596247"
/translation="MGEAFYIFFCLIGFLCSIIPSIWHWKYRNVAPLCLIFWVSSTNL
VYFINSIIWYNGSETSYRGDLYCDIVTKLILGSVTGELGATAAITHYLSKIMKPSYSF
LQSKITRRNQAIEDLLFSFGPPIMIMSLHYIVQPARYVIDGTSGCMPWTDRSWLAVAI
VLLWPPVFGSISAYYSVKVIISYIKKRNEFQTVLKDSKTSMTLSRFIRLIGLSSLIIT
IYLPLNIYLLIANIAEIIRSNIKYSWSHVHNWSSIIFYVSKSNMPFNRWLSPSSGIII
FIFFGMGSDAIVMYKEIAKKLYITHFFHFIQRKIFRKKTQDIKDAENYYNSYSFEKSL
GSCPPLFYNQTRDIRIVENGSLNECPSYPPIYTDHNNHNNNHNNHNHNNNHNHNHHN
HNHNHNHNHHNNHHNNHHNNHNITPSTYNQSHEHIFNTEYNKYRSHK"
CDS join(1359. .1887,1934. .2415)
/gene="ste3"
/note="alternatively spliced"
/codon_start=1
/evidence=experimental
/product="putative a-factor pheromone receptor Ste3b"
/protein_id="AAG38548.1"
/db_xref="GI:11596259"
/translation="MGEAFYIFFCLIGFLCSIIPSIWHWKYRNVAPLCLIFWVSSTNL
VYFINSIIWYNGSETSYRGDLYCDIVTKLILGSVTGELGATAAITHYLSKIMKPSYSF
LQSKITRRNQAIEDLLFSFGPPIMIMSLHYIVQPARYVIDGTSGCMPWTDRSWLAVAI
VLLWPPVFGSISAYYSVKVIISYIKKRNEFQTVLKDSKTSMTLSRFIRLIGLSSLIIT
IYLPLNIYLLIANIAEIIRSNIKYSWSHVHNWSSIIFYVSKSNMPFNRWLSPSSGIII
FIFFGMGSDAIVMYKEIAKKLYITHFFHFIQRKIFRKKTQDIKDAENYYNSYSFEKSL
GR"
mRNA complement(join(<7254. .8125,8261. .8431,8475. .>8889))
/product="Ydr124wp-like protein"
CDS complement(join(7254. .8125,8261. .8431,8475. .8889))
/note="similar to Saccharomyces cerevisiae ORF Ydr124wp"
/codon_start=1
/evidence=not_experimental
/product="Ydr124wp-like protein"
/protein_id="AAG38543.1"
/db_xref="GI:11596254"
/translation="MSDIMKSVLGMNESLIPKEKSISLNSTWDPQNFQKKKNMNKDHL
CIDVNTNDKSNEYDSWKTRYRSLVIGDTAEVIAFMESRFKQLQQSVCKIVAKAWIKVI
EPKKQTRYPYNRGDRSKPSWWPENIRHKEPDHLMKSERISLLITMLRCSKIPINRLEL
ATAEISTFIPPDKTSLLREIYHVARE EEKMRNGDIVKMGLRLLLLRQNHRSNRIKVIQ
MKLNRLIIYQSYRNIFVLIVLQMNKIPVYTKNHFSRIHVILSKWRINYIQNQNRLSSE
YLAYKKQKIKDTSNHTDAYNIEKDNNVNCQSLDNLSALKKNMNDSYHPMNFSDLKEDN
QNCYTNTSMFPLSSDQFFDSPYNFSLLDKNLLMHPKHLLHNYSSSSGSTSNAAISTYL
TTDLFSQKNHQQLSEMFSPNQFTSKYCQSYSNINHSCFPLHPDTFYNDGLPGTCFSMP
IKTSNFPDSTGTLQIKYGQDNTFKKTDESFVEDTN"
mRNA complement(join(<9655. .10008,10050. .10319,10361. .>10390))
/product="helix-loop-helix DNA binding protein"
CDS complement(join(9655. .10008,10050. .10319,10361. .10390))
/note="similar to Schizosaccharomyces pombe protein YAWC"
/codon_start=1
/evidence=not_experimental
/product="helix-loop-helix DNA binding protein"
/protein_id="AAG38544.1"
/db_xref="GI:11596255"
/translation="MSLYLILIILVDKVYRTEDALSKEFHGLSEITRYTGMNGEEEEE
EESGGIYSVESGEEILKSGVIRRDEEIHDAGGHKPIVGSEEWHRVRRSNHKVRVKREV
ERRRRETINESISELSRIVPGCEKNKGKILERTVSYIQQLKENEANNIEKWTLEKLLT
DQAISELSATNQKLKAECDRLWREVSVWKQAATCGINNDKDEKKDDPISNVRAEKMD"
mRNA complement(join(<10594. .10818,10890. .11744,11787. .11975,
12020. .12181,12434. .>12475))
/product="Yer049wp-like protein"
CDS complement(join(10594. .10818,10890. .11744,11787. .11975,
12020. .12181,12434. .12475))
/note="similar to Saccharomyces cerevisiae ORF Yer049wp"
/codon_start=1
/evidence=not_experimental
/product="Yer049wp-like protein"
/protein_id="AAG38547.1"
/db_xref="GI:11596258"
/translation="MFFLILRMNKVQHRSSDFITAISQEYHSARPYKHSVLTGIICEN
LLRKIRHEIIENLAFSEKETDIYKVFQTGDLENLSGLNENERNKLCSLIELRNALYLM
NFGYFISKVTGCGYLSGKKIDISVNMYGQGGHLLNHDDVIGSREVSFILYLIDPDETW
LPSYGGALRLYPTVRPCFPESEHEKAIFPVWNQLCFFQVQPGRSFHDVGKKCMKRING
GSVFQDGFHRPQPGEEGWKESEKSEEKVLSTLQNLHSNIQDYDEPRMEPRICSITFND
EFNEFLKGHLVYLKDWINPVYLNYEKMIQLRETFLDNSVLELPDFLESTVSSRIKDWI
IDIEKSSSTHLSLDTSSPWKLAKPLHKHRYLYIDTPDSSFNSPLISLLQNLFQSSFFS
SMDSFLHHFNSFNFYFTYSAAFVLPPSGHWNHGQNGGYCLYMEPNTITHSTLLYSTPS
WNNLHLILRDPGVLHFIKYVSLYSNFSRWDILSNMYFITS"
gene complement(<13024. .>15664)
/gene="ste20"
mRNA complement(join(<13024. .13095,13144. .13478,13527. .13617,
13683. .13915,13970. .14909,14960. .15039,15255. .15360,
15410. .>15664))
/gene="ste20"
/product="protein kinase Ste20"
CDS complement(join(13024. .13095,13144. .13478,13527. .13617,
13683. .13915,13970. .14909,14960. .15039,15255. .15360,
15410. .15664))
/gene="ste20"
/codon_start=1
/product="protein kinase Ste20"
/protein_id="AAG38545.1"
/db_xref="GI:11596256"
/translation="MAHNLSSQYFMNPGPAPAPPIKIPSYSSSNIYAGRCFDVSHTYK
GSKGVLKAGWANVKEEGIKSWIWSKKYLVLTDKVLRVMKNESSTSPGLSIPLREIMNV
SRAELKDYCFEILRDPGIGSGTKKSYFISLKSDNDLYSWMDEIYSRCPLMMVSNPTNF
THKVHVGFDPKSGGFTGLPEEWTRLLTASAITKEDYAKNPEAVIEVLQFYSDTQLREK
EEFGEPSTLGPDLYNIENQDIKSLEQNKMTMEKHYVNNGFNSQYPVSSTSKGAVPLRE
WLKKMDIKDSEKMVPQRVPPALPSNLPTRLNTDMASFPHESDSFLTCGSSNVDKGFGV
TAKKRQNVPLETLKHKQSQNNLAFISQAPTQPLPKIPTSSKNIPDIPTTFPETTMKQP
NATKLPLSLSNARAMKQEPPSDEKVNDLTGSQVMERLRNVVSTNDPNPLYTKIKKVGQ
GASGSVYVARVAIKQMDIAHHPRRELIVNEILVMKESHHPNIVNFLDAYLAKNSELWV
IMEYMEGGALTDIIDNNTLTEDQIAAISLEGLQHLHPTKIFHRDIKSDNVLLDFHGRV
KITDFGFCAKLTEQKNKRATMVGTPYWMAPEVVKQKEYGAKVDIWSLGIMAIEMIETE
PPYLNEEPLKALYLIATNGTPTLKKPDKLSKELKSFLAVCLCVDVKSRATAEELLEHD
FLKKSCSLSNLSSLLAFKSGK"
gene complement(<15925. .>18086)
/gene="sca1"
mRNA complement(join(<15925. .16243,16294. .16567,16643. .16829,
16868. .16923,17004. .17098,17185. .17443,17514. .17863,
17926. .>18086))
/gene="sca1"
/product="putative sucrose carrier Sca1"
CDS complement(join(15925. .16243,16294. .16567,16643. .16829,
16868. .16923,17004. .17098,17185. .17443,17514. .17863,
17926. .18086))
/gene="sca1"
/codon_start=1
/product="putative sucrose carrier Sca1"
/protein_id="AAG38546.1"
/db_xref="GI:11596257"
/translation="MIGLYAHVNASKDHFDDVINASPPLEEPQEPPLRGTLYLLFLTL
VLGGLQLTCKSRKIFILYYIREGTPYLLSLGLTKHYTSFVWIAAPLTGIIIQPLIGFF
SDRSRSKFGRRRPFLVVGTLCVVISLIGLAYFKDIAFFFFSGHEKVAKNMTIIVSIII
VYLLDFSINIVQASSRALIIDMVPMVQQDLANAWASRMIGIFNVVGYLNGYLNLPKIA
PALGNTEFKALSVIGSIVLIVCMAVTFFVAKEKRYFMESLFNFFKTIFSTIITLPSEI
LWICIIQFFAWIELISMVFKVSDSFLCKIFFGTTYVAEIYAEKYCNKYLKYPDHLYTK
EKEYLLIEEATRYGSFAFCAFFYYFFIRNIFLSFIFFMDFWSFYFYFMHAFNHTSKIN
HFFFPLFVSGKKLFKYIFRLKFYNILQNINIQFLWTFGHFIFTLCMLSTILVKSITQA
TIIIALCGLSWAITLWAPFSLIAIQLSSNNQLHRSGTILGVHNIFITVPQVLSIIMVG
IIFKLTGYKKFEDIIKCNDNMDYSFIWIFQISGISALIAMYLSFKIYTNDPSYERNKI
"
gene <18200. .>20207
/gene="nhx1"

Alignment Scores:

Pred. No.:	6.6e+03	Length:	32083
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0

```
Query Match:            100.00%       Indels:        0
DB:                     15            Gaps:          0

US-09-858-332G-28 (1-6) x AF309805 (1-32083)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       2608 CATAATCACAACCATAAT 2625

RESULT 410
AX571959
LOCUS       AX571959               32816 bp    DNA     linear   PAT 29-NOV-2002
DEFINITION  Sequence 3 from Patent WO0246407.
ACCESSION   AX571959
VERSION     AX571959.1  GI:26004050
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Chaturvedi,K., Wei,M.H., Ketchum,K.A., Difrancesco,V. and
            Beasley,E.M.
  TITLE     Isolated human transporter proteins, nucleic acid molecules
            encoding human transporter proteins, and uses thereof
  JOURNAL   Patent: WO 0246407-A 3 13-JUN-2002;
            PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1. .32816
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:              6.75e+03      Length:        32816
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-28 (1-6) x AX571959 (1-32816)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       9942 CATAATCATAACCATAAT 9959

RESULT 411
AC142293/c
LOCUS       AC142293               32823 bp    DNA     linear   PRI 13-MAY-2005
DEFINITION  Homo sapiens fosmid clone XXFOS-8806F2 from 4, complete sequence.
ACCESSION   AC142293
VERSION     AC142293.2  GI:29469646
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 32823)
  AUTHORS   Bielicki,L. and Levy,A.
  TITLE     The sequence of Homo sapiens fosmid clone XXFOS-8806F2
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 32823)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2003) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3  (bases 1 to 32823)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-2003) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 32823)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-APR-2003) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 32823)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUN-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 32823)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Apr 2, 2003 this sequence version replaced gi:29294412.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_AA008806F02
            --------------.

            NOTICE:

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            SOURCE INFORMATION:
            This fosmid was derived from a human female fosmid library.
            VECTOR: pepifos5

            NEIGHBORING SEQUENCE INFORMATION:
            This sequence is not the entire insert of the clone. This clone is
            overlapped by AC116645.
FEATURES             Location/Qualifiers
     source          1. .32823
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /clone="XXFOS-8806F2"
                     /clone_lib="XXFOS"
ORIGIN

Alignment Scores:
Pred. No.:              6.75e+03      Length:        32823
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     8             Gaps:          0

US-09-858-332G-28 (1-6) x AC142293 (1-32823)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      29294 CATAATCATAATCATAAT 29277

RESULT 412
```

```
AC114261
LOCUS       AC114261                33651 bp    DNA     linear   INV 12-MAR-2003
DEFINITION  Dictyostelium discoideum chromosome 2 map 126059-159710 strain AX4,
            complete sequence.
ACCESSION   AC114261 AC115582 AC116031
VERSION     AC114261.2  GI:28829373
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 33651)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 33651)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 33651)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 33651)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On or before Mar 4, 2003 this sequence version replaced
            gi:19697368, gi:19569921, gi:19424365.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES             Location/Qualifiers
     source          1. .33651
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /chromosome="2"
                     /map="126059-159710"
     CDS             complement(join(3028. .8216,8381. .8855))
                     /note="GeneID exon scores (in order of location ranges):
                     473.34, 31.97 - GSCJ_ID dd_01993"
                     /codon_start=1
                     /product="similar to Homo sapiens (Human). HERC2 protein"
                     /protein_id="AAL88727.2"
                     /db_xref="GI:28829374"
                     /translation="MSLFFKSIKSKIEETLNHTPQVNLFNNSNNGGNNSNNGGNNSTP
                     TLTKTPSTTNFPYNDYKNYTYKYVYDNTDNYNFNNNNNNNNNNNNNNNNTKENNFKS
                     IRESASIYKNNEEINNDDNNTIINKKKIEEEEDETNFDTSLIPKLGHEDKDELESDHY
                     ITIWGNGIRNYKNCNMSSIGFSEMYNQQKIVLISTGAIHSSFITDQGNLYTFGDGLLG
                     KLGHGNLESYSSPKLVEFFQTKPSLRVVSIANGGKHSLALTICGKVYSWGGNDSCQLG
                     LGNGSSSSYYSLPQLINFNISYNNNNNNNNNNSTNNNNNNNNDGAQQQFSLSQNSSSN
                     NNKIKIVKLSCGTNRSAVITDKGLLFTFGRGDHGRLGLGEQSLLMQSTPKLVQSLQGH
                     FIIDMSSGGGHSLVLTNKGLVFSFGRNDQGQLGIGCFSNQQTIPKQIQWFQTSQQQQQ
                     QPINIIKVSAGGYHSIAISDNNDTYAWGRSDYGVLGTSIDVQGDRNLPVLINSNSVVV
                     DNVKFIDVSSGFQHNVAMSIDGSLYSWGCNAGQRLGFKSDENQLTPKKCFINSKDSKP
                     LLFTCGEIITIIVEKRFKEPIIKTELQLPSSTQSQQQTELSLPSSVNSSSDSNFNKSI
                     IDNSSTTTIIKQTNNNIKSTTTTIITRPSSPTPSFLSLANEGKDLLNSTQSNFIEWMT
                     LISKSKGLLNLIQMKATSSLPNLGLSIEKKNKDFDDINRQLKLLVNSNLSLSPLPDNI
                     QSEIDRLVLSRNQLGLGLVSEYQTSIDYWLEIKQQKDLTFLNLSKLLNLSTITTTTTS
                     NTTTTTTATTTTTTTTTTNLINKKVLVHQDEKQQQREKSETELEEEQDEEEEDSEIKN
                     GTLTAIEIQQQSFYLVERLNQGLGNFCNNIMDSTTMVQPLTTATASSTTATATATTTT
                     ANTTNSSGIINKNFGESLRLSMEEIQLIVLETQKYAIKYHSLLSQQLQQCDDLINIHR
                     GKLVEVQDFLNKSKLAQSLLEQREKLKSDYRFCKRNSIELSKKIEVIELDYDQNLCDD
                     DDNNHENNSVNNNSNNNNNNNNNNNNNNNNNNNNNNNNNIDNNINSNSINDSSNNNNNNN
                     NIEKLKNELIKLKKDEQKLLSNQQEINELIIEILDKYVPEFKIKLRANDKISSRIKDT
                     GLLVTDRKFQHYDIIKTLSTHPHNVYLANFNDQLVVLKEFGIGDAFGKQIFERQVSLL
                     KQMNHKCIMPIQAVFYDRNAFMQMEYISGGNLLDWCRNNNNDNENNKRRQPWEIQKIF
                     QQIIQAIAYMHSNCIIHRDLKLENVLIRQDGTPVISDFDLSKDISANVNVTVFNINGG
                     TELYKAPEMKEQNVKGSYSTDIWAFGVMLYKCIFQFVNNNNNGNNGKNNNSDGNENDN
                     NNNNNFLIIREPFLLPEENNLPLPANHSDQRLISLLSSIFQRNPKLRPTAHQIAVHPY
                     FVTSLVEDLLSSRTLIDCREKIAAFRAHISSLSEMAEEMSESLQLTVRREHLVLDFFQ
                     FFFKKIESNKLFCRLEVSFQGEKGLDLGGLSSEMYSLLFSDNQIIDNSNNSNNSVSYS
                     IPKNSLFSKKFNLFENSGTESPFYLLNSNDLFNNNNNNNEENNNNNNNNNNNNNNNNN
                     NNNNNNNNNNNNNNNNNEENPLLILKNEFTIFKILGKIFLKSIIDGKPIPDCFPTSFFK
                     YLLGVKVNLRDLEIYDPQLAQSFKKVLVLDNIEEYLSTTFEGLIEGGESIPVTDLNKE
                     EFIQRNIERVLVGCRQSKLEAFKSGFMSIDSLNAHFALFSPTELQLLMCGNTLVDSSV
                     LQKNFKFIGFPDTSSTPKDFRRAVDEMNQDEIRLFLRFVTGMVALPLSGFEKSISIIQ
                     VPLSQKLPCAHTCSYQLDLPDYNDFDTTKKKLIKMLEYVDGFAFI"
     CDS             join(9500. .9672,9905. .11153)
                     /note="GeneID exon scores (in order of location ranges):
                     -2.72, 115.02 - GSCJ_ID dd_01994"
                     /pseudo
                     /codon_start=1
     CDS             join(11756. .11825,11932. .14426,14517. .18431)
                     /note="GeneID exon scores (in order of location ranges):
                     2.47, 246.65, 349.58 - GSCJ_ID dd_01995"
                     /codon_start=1
                     /product="similar to Dictyostelium discoideum (Slime
                     mold). Non-receptor tyrosine kinase spore lysis A (EC
                     2.7.1.112) (Tyrosine- protein kinase 1)"
                     /protein_id="AAO51915.1"
                     /db_xref="GI:28829375"
                     /translation="MSNNNYNKIQSPSSGKRRFLQEIPKGEEDELYSEHFIEIDKLIK
                     QEQEKEENGGFSLPKWPASSTINNNSTQNLFSNFMSDSSSSINSDINSNYTRDNNGMF
                     IVRTDLNTDIPSSHKSTIYSKQQQSPRHQDINSNHYSMTYDSSLLSVSGSNASSNCNT
                     SASSSVNSCNNSIGNLSKVFNLNSSTKSLNNNNNNNNNNNNNMETISGFNNIMKIDNL
                     INHDTNNNNNNNNNMITSPSLTTNTTTTTTMTSNNSTPIRSPSPPKTPQHQSVLISPMT
                     PYDVYNSYKFDKLKTTPTSPTSPTKSTRFKEPLLPRSNNNFNNNNNKNSTMSLSGMMS
                     GTFSTSHLSSWVWVKNKHSIVIIGPTGIGKSFLLNTLLHSTFTPSLNSIRPILGNHRL
                     SLKPQHDTTSTSTTSTTNTISTDHIMSNNEYFLLDKLKTKLMNISKSQIDQKMILDQS
                     FYQSINWDKERDNEFERVYTTSQYLTKKIPQVNDSHAFILPQGRLGPNQQFKFTYGAI
                     CELNVEFLQESEVRVILWELHQYSEDKTIPMNPNRLQSLQWVYSSVVDGATFDYSNQS
                     SSSSTSSINSGTSGSNRMVRVVEDINSYWDIPIKEKFKKLFGKRIQFKGLGKDHNIDR
                     IYVREKVKEVLSVYGCILNLLHIKLPSSVLFNKREITMISDDSKRGNISKFTITRILD
                     DADQILLSIDYRGLSADLADLLLASKFTQKFSADDGSIQFLELGEHRQFQLQQQLQQK
                     LQQQQVSSPQPQPQQQQQQQQVQQCQSQQSPQQCQSQQSPQQCQSQQSPQQCQSQQSP
                     LQQQQQQQQSPQQLQSPQQLPDIHSFVINQLRKLFKDARNISNIHRILDSFKIIKIKP
                     LLFSSIHLNKSVPPEYKLKGYSSIRQLKEISNVPVLIRLMHVSNNSTWTTQHNLLDKF
                     KQGIAKKPSILNHLVEQQQQQYQQQDALNSTGSKPIPTDKELNDLQEQLTNSFKDTLN
                     VNNNQNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNYNNNNYNNNNYNNYNYNNNNVIT
                     KNNNNVINNNNNYNNNINNNNWRNNYNRGVINPPSGIDRITTSFRTRFAGWKRKLLAN
                     VSSEASESLFQNKSNECEQMARMEWIKFKMVSPDAVVNINDFCNLVYSTFGGTNAKPD
                     TLSNVNGNNPVILLCKTLLQWTPNFWRENIVERLPTLQEEISSFSKTLLFETNKLLLN
                     RITLPNDEKNYLMKVLHYCQAATDRNISSKISGVQTLFNQNMNLIIYNTCLDTFCSEV
                     TNIKFQQSSVNYHLIMESCFSIIPYAVSTKIQQKLIEILRQGLSYLDSIEKHIEYTIQ
                     YLLNSPFRIKRQQQLKPSIDVSDSVIENYVNLVNPITKQSPLKSQSQQTTTTTTTTTT
                     TTTTTTTSSPSNSPKLSTDEMDIETPNKKIKSDNSVNNINNTTETTPTETSPNNSSNV
                     IPTPMIINNQTTTTTTTTSSTISNNNNTNNNNNNNKPMFDNGSCNYIPKQILPPQTAD
                     SLIQICVGKIVQNIHSLTSVKGTTIPEEILQKIITLLISQDKIKGGSNDPITNQLDDS
                     LLARLLSPFMQSLDLEGAKFLSTISLKTIGSTCSQLKKLSLANCINIPSDALNSISMS
                     CKNLEVIILKGCYQLSNPGIVSLARGCPNLYVVDLSGCMKITDFAIHELLQNCKQLHT
                     LDLRKCVNLTDGAFQSFNITTLANIDLLECNYISDQTIFNICSTSRNLLSIKLSGKGI
                     TDQSLKKISENCQSLTNLDLVLCENITDQGVQLLGKNCLKLSSINLFSSKNLTSSVFD
                     ETINNNNNNNNNVNNNNNNNNNIVNNNNNNNNNLNNNNNNNNNNMIINIFNQQSWSSLT
                     SLNLNRCITINDTSILTITNQSPLLETLILAMCTDISDESVITIAQRLKNLKNIDLTK
                     CTQISDRGVIEIAKQCKQNLNRLILVSCTQVTDASIIEVANQCSSLIHLDLSQCEKIT
                     DQSLLKVSQGLRQLRILCMEECIITDVGVSSLGEISEGYGCQYLEVIKFGYCRFISDS
                     SLIKLAFGCPFVSNLDLSQCSNLITPRAIRSAIKAWPRLHTLRLRGYQSLTNESIVES
                     TPLKLKTVNLSWCANMEDSALIGFLKQCTAIETLDISKCPKITDNSLESILDSCPSIR
                     VINVYGCKEISSFTVQKLSSLGKAIYR"
     CDS             complement(join(18692. .19218,19339. .19525,19609. .19653))
                     /note="GeneID exon scores (in order of location ranges):
```

```
                34.75, 0.48, -3.97 - GSCJ_ID dd_01996"
                /codon_start=1
                /product="similar to Dictyostelium discoideum (Slime
                mold). Developmental protein DG1037 (Fragment)"
                /protein_id="AAO51916.1"
                /db_xref="GI:28829376"
                /translation="MVNQLSSKESEIFLISNRFDEKEIKIFENIFNLIDLESKGYICK
                EDFDQYKSNKTGLNSFNIEFEKFLIFNNSTITFWISLNITTPPIENESKSISSSYSRL
                SFVPTSVAQQQISQLFNNNNNNNNNNNNNNNNNNNSNSPQFSSSNSTPTIQFNNSPP
                SSPSILTSTHLLNSHSSSSSNLVPITTTIAESTLIETIGKGSLNNIGLNNNNNNQNNQ
                NNNNNQNNNNNNTDSDNENKRTSRMGLGGHRGKS"
     CDS        20046. .23120
                /note="GeneID exon scores (in order of location ranges):
                116.33 - GSCJ_ID dd_01997"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAO51917.1"
                /db_xref="GI:28829377"
                /translation="MNDEYNEKLFWKVFKNKYLLNYILKNKCCHKSKKFKTIISINSM
                LNHKQYQLLICKLKANENLYYGIDFESRFPKILENIGEIDDFEAFKEIHSLLFLKGYE
                HHGNNYSWYMKLLTRFDLGPKFEIYLDLHLKYYQQELKQYILNNMDEVFYYFDSKCIN
                KILNCFQQQSISNKPFTLSSFERNHLLIKFLNRSNPKELMKKLKIISKIPISPTSQLK
                RIDENLRLSFGLYLEYLKQLKQFGLQNIPTFLLCQKLSGGEFNSEIKRYIMNDKGNLE
                YSSLSVYSILLYFKMYYNSYDYRSIPRSIQTQIDIIIENSKHRRCTPYQSTKQLFQLL
                ASETKSIALYTLYLKEYDEPIETIEINEKILLYIMIKTYSINILKFLLKNHIEIESEK
                NKLIDDFDLYDLGKIFKNSKFDKRNIIKFFKVFSSQKSIIIKNFDQDKELYPISFNQL
                CMSKWKEIISNLIDKDMIVTIKDKETIDWLSDTKIGGCSYITIRDALNSKLKQLKLQF
                DTLELLDYARLKLNQIIQNFDSNQEICNIHILRTHDDIVSLINRLNESNFKNNEDLFK
                YYMVQIIQFGSRNSIKYQRLFCDNFDELRYSGTKSLDEYINTIKFVGERFIESSNCCE
                SKPEYRKNEKPRNYKIPTQQTCQIINLSKIKVNVSHFPNATSYLLQKFIQLLDIDGVQ
                FILDNLNGNGVGIDEYSFIDTSGLKKYICGYLYKDLMVFIQFIIDKFNQHQYNISLSN
                LSFSLNFLFSNLLRIKNLTIDQIKSAYQLISNYVEINCSDIHSHYYLNTLSPLIGSLS
                VRSFSIERFCTYETDAKVYTFGYIFKNRDNNDLHQFIDSTIMDILKKKPFKRLISFHS
                LLVHMLDNNNPRINLFLYYFKMQLEFIKTNEIKEPFLHYKLFKLIIKTFDLESFIKLD
                NLLQQYNFYFKKDEDSDEEECEYKNSKKKPKTSTSNMVEWFGVNIYKDYSDYPDIINY
                LYSKYSLQLEPYLEYYAFDRIKQKTFEQYCTKIPHGRRIINKKSGAERVFGI"
     CDS        complement(join(23315. .25933,26024. .26284))
                /note="GeneID exon scores (in order of location ranges):
                272.84, 16.73 - GSCJ_ID dd_01989"
                /codon_start=1
                /product="similar to Dictyostelium discoideum (Slime
                mold). Protovillin (100 kDa actin-binding protein)"
                /protein_id="AAO51918.1"
                /db_xref="GI:28829378"
                /translation="MEPPLELPTQRKRVIPSKFGILKRNAEIEAEKNRENLQQSSCFS
                HINEIGKEIGLEIWKIIDDSTIQKVPKVNHSTFETNKSYLLLMGQFYDGNMNIKTYNI
                HFWIGELLINSQETINFCNDRIEELERIIKYNQKQFDSEQFYPEPILYREFQGKEGDI
                FMSYFKSYGGPRYVAPLKLTSASAAIATAAKQYKLFHLKGRRNIRVKQVDISSKSLNS
                GDVFVLDCEDFIYQWNGSESSRLEKGKGLDLTIRLRDEKSAKAKIIVMDENDTDKDHP
                EFWKRLGGCKDDVQKAEQGGDDFAYEKKSVEQIKLYQVENLNYEVHLHLIDPIGDVYS
                TTQLNAEFCYILDCETELYVWLGKASANDQRTVAMANAMDLLHEDNRPSWTPIIKMTQ
                GSENTLFKDKFKKGSWGEYVNDNFEKKPITGKGVAAKAVQEKINVDALHNPEKYQLSK
                EERKSTIPTLHHVDDKHRGELKIWHVRNRNKFEISQSEFGLFYNQSCYLVLFTLFAAD
                GSNNSILYYWQGRFSSSEDKGAAALLAKDVGKELHRSCIHVRTVQNKEPNHFLEHFQG
                RMVVFKGSRPNATTEVSLENLSSSLQGLYHVRGTEPINIHSIQVEKAISSLDSNDSFI

Alignment Scores:
Pred. No.:              6.91e+03        Length:         33651
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-858-332G-28 (1-6) x AC114261 (1-33651)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      14107 CACAACCACAACCACAAC 14124

RESULT 413
CET05C12/c
LOCUS       CET05C12               33780 bp    DNA     linear   INV 09-AUG-2005
DEFINITION  Caenorhabditis elegans Cosmid T05C12, complete sequence.
ACCESSION   Z66500
VERSION     Z66500.1  GI:1041376
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 33780)
  AUTHORS   .
  CONSRTM   C. elegans Sequencing Consortium
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology
  JOURNAL   Science 282 (5396), 2012-2018 (1998)
   PUBMED   9851916
REFERENCE   2  (bases 1 to 33780)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-1995) Nematode Sequencing Project, Sanger
            Institute, Hinxton, Cambridge CB10 1SA, England and Department of
            Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
            worm@sanger.ac.uk
COMMENT     Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT:  This sequence is NOT necessarily the entire insert of
            the specified clone.  It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
            For a graphical representation of this sequence and its analysis
            see:- http://www.wormbase.org/perl/ace/elegans/seq/sequence?
            name=T05C12;class=Sequence
            IMPORTANT: This sequence is not the entire insert of clone T05C12.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone T05C12 is at 1 in this sequence. The
            true right end of clone T05C12 is at 4883 in
            sequence Z49968.
            The true left end of clone M110 is at 33681 in this sequence. The
            true right end of clone F10B5 is at 15182 in this sequence. The
            start of this sequence (1. .101) overlaps with the end of sequence
            Z48334.
            The end of this sequence (33681. .33780) overlaps with the start of
            sequence Z49968.
FEATURES             Location/Qualifiers
     source          1. .33780
                     /organism="Caenorhabditis elegans"
                     /mol_type="genomic DNA"
                     /strain="Bristol N2"
                     /db_xref="taxon:6239"
                     /chromosome="II"
                     /clone="T05C12"
     gene            join(Z48334.1:23435. .23502,Z48334.1:23558. .24167,
                     Z48334.1:24214. .24449,Z48334.1:24497. .24610,
                     Z48334.1:24661. .25018,Z48334.1:25064. .25883,
                     Z48334.1:25931. .26489,Z48334.1:26532. .26743,
                     Z48334.1:26790. .27477,Z48334.1:27526. .28249,
                     Z48334.1:28294. .28751,Z48334.1:28797. .28902,
                     Z48334.1:28954. .29132,102. .114,161. .358)
                     /gene="rrf-3"
                     /locus_tag="F10B5.7"
     CDS             join(Z48334.1:23435. .23502,Z48334.1:23558. .24167,
                     Z48334.1:24214. .24449,Z48334.1:24497. .24610,
                     Z48334.1:24661. .25018,Z48334.1:25064. .25883,
                     Z48334.1:25931. .26489,Z48334.1:26532. .26743,
                     Z48334.1:26790. .27477,Z48334.1:27526. .28249,
                     Z48334.1:28294. .28751,Z48334.1:28797. .28902,
```

Z48334.1:28954. .29132,102. .114,161. .358)
/gene="rrf-3"
/locus_tag="F10B5.7"
/standard_name="F10B5.7"
/note="C. elegans RRF-3 protein; contains similarity to Pfam domain PF05183 (RNA dependent RNA polymerase)"
/codon_start=1
/product="Hypothetical protein F10B5.7"
/protein_id="CAA91312.1"
/db_xref="GI:3879453"
/db_xref="InterPro:IPR007855"
/db_xref="UniProt/TrEMBL:Q19285"
/translation="MLPFDNDDSSDDATTSVRPKHPRGVPQSQSTFPRGRSNFSSGTL
PNRKTECTPVNTLTIGHSNKMLLTTFRMDRNSKSKSEVDVQEQPVHSSSSAFPGNHLN
NFSYPVNRGYLRDYLLQSQRPSTSKPVDCSVLKRHSLPSTHILYEKTKHRGGVNIEEQ
EKLVRMLWAAAEESETVAKTRQFSKKQAIELNFDAKLIGSMNNDCFGYCRAHMENIKD
VLKTHLKLSKVDEVNWIKVGMVPRAAYEDKSYVIDAHLVLTPNGEVEDENELFSEFAS
SFTSRITGMLHDQVFLEVPKMHTLFTKITPQHMDINISAIAIGNCPNSGLFLVRGDFI
SQENTVCSVKLQSHHNADASRENSSFKVAGSNKYLSYARFEHDKRLAVVYFGVRLAEF
ADDGLDHAGFRLNLYYNLFVRIVVDMSHETTNSIYIQMKNPPHLWEGIPKNTIFHPSK
SKVLNMETCTEWTRVLSWPGDAEGRGVGCTSEAFSQSSWIRLTMRKDDDNDSVSSTQL
MDIVTRLSARSKAKVMFGSIFSIRRKLAPSPAFHSLGSFRANYALQALITRGSVFTDQ
LFDATDENIPSSDNDNDEDDDDDVDDTKKPMELVHEPLFLKLVRRGMKECSQATEETL
EQLLNAFDERRQIDVVTAFTTMYQSRKIQYERLLKGESLQDVGLAKPLPKNCVSVAKV
IVTPSRILLMAPEVMMVNRVVRRFGPDYALRCVFRDDNLGRLAIRDFSINNIDHMSNI
VTEGIYLTLKNGIQVADRVYSFLGWSNSQMRDQGCYLYAPRVNALTGEVTGTVEDIRV
WMGDFRDAISVPKMMSRMGQCFTQAQPTVYSSVKNIHIVENIQVRLERHHWIVEPDIE
GGVENKYCFSDGCGRISIKLATHISKILQLKEVPACFQVRFKGFKGILVIDPTIDDII
NMPKVIFRKSQQKFGEGGGELQDEYLEVVKYAMPSPVCLNRPFITILDQVSEKQSASS
HRRITNRVHYYLERELCSLSNMLINENQAAEELVNRTNLAIDWNAASKRAGFELSVDP
LIRDMLFSIYRYNIIHHISKAKIFLPPSLGRSMYGVVDETGLLQYGQVFIQYSPSIRQ
TSNRPILKTGKVLITKNPCHVPGDVRVFDAVWQPALAHLVDVVVFPQHGPRPHPDEMA
GSDLDGDEYSIIWDQEMLLDYNEEAMVFPSSSAAEEDKEPTTDDMVEFFLRYLQQDSI
GRMSHAHLAYADLHGLFHENCHAIALKCAVAVDFPKSGVPAEPLSSFEQCEMTPDYMM
SGGKPMYYSTRLNGQLHRKARKVEEVLEEFETRGSVFEREYDKLICPEDVDVFFGNEI
KLVQTLTLRDEYVDRMQQLLDEYGIEDEASVVSGHAASIKRLAGMERDDYSFYHTDKV
VELRYEKLYAVFRAKFFEEFGGEEINIENDGKNTRLKCTKAMHEKIRQWYFVAYVQPK
INKAGRCIGQSLPWVAWDALCDLRRQLMLDKNDAVLRGKYPIAARLEEEIENSIERQF
DKFLKLKDLIESHKDALFLRRYVYFYGDQIIKMLFILKVWLERENVLPSSVLSIWQLG
RLLIRLGLGDLLGNPTIDYEKSLLMPTTMFQQWISKKEDADEAPILRNFDMGTMMLEF
LRYLASQSFASAESISLRVFYEKDIVEPILTKSAQWMPLHLIAYRTFHSIAVSGRFDA
LHLDDEDAVDQITESKDPILVNESLFSSRNYNDDYPISRSRILQSLKDWSGVKEIIPR
EITGTRKSDMIYVTSVGTVLARQRLARLLLLSGETIRDAIANNVVPNEVRDEFL"

gene complement(join(528. .1073,1173. .1350,1716. .1813, 2088. .2204))
/locus_tag="T05C12.11"

CDS complement(join(528. .1073,1173. .1350,1716. .1813, 2088. .2204))
/locus_tag="T05C12.11"
/standard_name="T05C12.11"
/note="contains similarity to Saccharomyces cerevisiae anchorage subunit of a-agglutinin; SGD:YNR044W"
/codon_start=1
/product="Hypothetical protein T05C12.11"
/protein_id="CAE17939.1"
/db_xref="GI:33300442"
/db_xref="UniProt/TrEMBL:Q7YWU2"
/translation="MIILLILLLFNVVLSVSFVLDYNNNFEIKKEYFQTDVIQEKLIRV
NEWHDIPVKGTEDEDALIDFHLFATCEKFSVLFMTEPNESTRFEFSYYDVTLKKIDVQ
SIGNAVQVPVSPKMLGTCRENNRKILRIKIRTSNDTITFKVDDQIIRRPFDSTQNNIK
ILIPQKVACCLIRALFLNNSDLDAYTSVSTRFFTSSKSSYKRILEVFGSTSSSLVLTT
TPSTSTSRTTTKNDTNKWSKFTLGKNPTVSSSTSEPKMESSWTHAVNKEWSITFVLCS
VIVCTIMLFLAVFGTFTVFYLMTKPRGGKSPFAVDA"

gene complement(join(4122. .4259,4309. .4644,4689. .4845, 4890. .5116,5314. .5385))
/gene="dylt-3"
/locus_tag="T05C12.3"

CDS complement(join(4122. .4259,4309. .4644,4689. .4845, 4890. .5116,5314. .5385))
/gene="dylt-3"
/locus_tag="T05C12.3"
/standard_name="T05C12.3"
/note="contains similarity to Pfam domain PF00106 (Alcohol/other dehydrogenases, short chain type)"
/codon_start=1
/product="Hypothetical protein T05C12.3"
/protein_id="CAA91310.1"
/db_xref="GI:3879451"
/db_xref="GOA:Q22230"
/db_xref="HSSP:1NXQ"
/db_xref="InterPro:IPR002198"
/db_xref="InterPro:IPR002347"
/db_xref="UniProt/Swiss-Prot:Q22230"
/translation="MACKNPKKFFPICNSPILRDGALKGKVALVTGGGTGIGKAIATT
FAHLGASVAIAARRMEKLEQTAEEIMKTTGGICEPFRMDIKDPGMVSDTFDKIDKKFG
KHPDILVNNAAGNFIMATERLSPNAHGTIIDIVLKGTMNVTTELGKRCIQSKTGASVT
SITAAYARSGAPFIVPSAVSKAGVEIMTKSLATEWSKYGLRFNAVSPGPIPTKGAWGR
LFSGEMGDVAEKMKELNPEGRSGTPEEVANLVAFISSDHMSFMNGVIIDLDGGQQHFN
HGSHMGDFLHTWDQDTWGDVENVIRGRTGKEKP"

gene join(5551. .5559,5605. .5793,5845. .6467,6517. .6673)
/locus_tag="T05C12.1"

CDS join(5551. .5559,5605. .5793,5845. .6467,6517. .6673)
/locus_tag="T05C12.1"
/standard_name="T05C12.1"
/note="contains similarity to Pfam domain PF00069 (Eukaryotic protein kinase domain)"
/codon_start=1
/product="Hypothetical protein T05C12.1"
/protein_id="CAA91303.1"
/db_xref="GI:3879444"
/db_xref="GOA:Q22223"
/db_xref="HSSP:1CKI"
/db_xref="InterPro:IPR000719"
/db_xref="UniProt/TrEMBL:Q22223"
/translation="MLNRNRNVNKQNGHQKQNGAQKSELRGKNAPPYLERETCLGRND
IFVVEKMIAGGGFGQVYRARNTETQEEVAVKVERATTNDSQRMILESKVLDDMLGSKH
FPNVYYIGPYHSYNFIVMQMLGKNIGDIRKMMPNKKISILSSVRIGIQIIEALSLLHS
KGWLHRDLKPTNCCLGLDEKRKTVYLVDFGMSRKFRNDNGSLRESRTYCGFRGTTRYC
SYRMHDRREQGPVDDLWCLYYSLGELIEGCLPWRDIESADEMAHVKKILKHEDIFHSM
PSKFASFGRNLRRLRPANTPDYTKFQNILAYCVKFVDDNAEFEWDVMDY"

gene complement(join(6783. .6905,6950. .7062,7158. .7350, 7394. .7590,7639. .7818,7877. .8008,8051. .8162,8207. .8309, 8355. .8409,8514. .8583,8627. .8761,8952. .9041))
/gene="acr-14"
/locus_tag="T05C12.2"

CDS complement(join(6783. .6905,6950. .7062,7158. .7350, 7394. .7590,7639. .7818,7877. .8008,8051. .8162,8207. .8309, 8355. .8409,8514. .8583,8627. .8761,8952. .9041))
/gene="acr-14"
/locus_tag="T05C12.2"
/standard_name="T05C12.2"
/note="C. elegans ACR-14 protein; contains similarity to Pfam domains PF02931 (Neurotransmitter-gated ion-channel ligand binding domain), PF02932 (Neurotransmitter-gated ion-channel transmembrane region)"
/codon_start=1

Alignment Scores:

Pred. No.:	6.94e+03	Length:	33780
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-28 (1-6) x CET05C12 (1-33780)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      10579 CACAATCACAACCACAAC 10562
```

RESULT 414
AC149414/c
LOCUS AC149414 35913 bp DNA linear PLN 27-OCT-2004
DEFINITION Phakopsora pachyrhizi clone JGIAFNA-77J10, complete sequence.
ACCESSION AC149414

```
VERSION     AC149414.1  GI:48056645
KEYWORDS    HTG.
SOURCE      Phakopsora pachyrhizi
  ORGANISM  Phakopsora pachyrhizi
            Eukaryota; Fungi; Basidiomycota; Urediniomycetes;
            Urediniomycetidae; Uredinales; Phakopsoraceae; Phakopsora.
REFERENCE   1  (bases 1 to 35913)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 35913)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUN-2004) Production Genomics Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA
            94598-1698, USA
REFERENCE   3  (bases 1 to 35913)
  AUTHORS   Stanford Human Genome Center.
  CONSRTM   DOE Joint Genome Institute
  TITLE     Direct Submission
  JOURNAL   Submitted (27-OCT-2004) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 100% of Sequence;
            Estimated Total Number of Errors is 0.
FEATURES             Location/Qualifiers
     source          1. .35913
                     /organism="Phakopsora pachyrhizi"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:170000"
                     /clone="JGIAFNA-77J10"
ORIGIN

Alignment Scores:
Pred. No.:              7.36e+03        Length:         35913
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x AC149414 (1-35913)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       8450 CACAATCACAATCACAAT 8433

RESULT 415
AC149320/c
LOCUS       AC149320               36340 bp    DNA     linear   PLN 09-MAR-2005
DEFINITION  Phakopsora pachyrhizi clone JGIAFNA-101K18, complete sequence.
ACCESSION   AC149320
VERSION     AC149320.2  GI:60650314
KEYWORDS    HTG.
SOURCE      Phakopsora pachyrhizi
  ORGANISM  Phakopsora pachyrhizi
            Eukaryota; Fungi; Basidiomycota; Urediniomycetes;
            Urediniomycetidae; Uredinales; Phakopsoraceae; Phakopsora.
REFERENCE   1  (bases 1 to 36340)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 36340)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUN-2004) Production Genomics Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA
            94598-1698, USA
REFERENCE   3  (bases 1 to 36340)
  AUTHORS   Stanford Human Genome Center.
  CONSRTM   DOE Joint Genome Institute
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-2005) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Mar 9, 2005 this sequence version replaced gi:48055860.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 100% of Sequence;
            Estimated Total Number of Errors is 0.
FEATURES             Location/Qualifiers
     source          1. .36340
                     /organism="Phakopsora pachyrhizi"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:170000"
                     /clone="JGIAFNA-101K18"
ORIGIN

Alignment Scores:
Pred. No.:              7.44e+03        Length:         36340
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x AC149320 (1-36340)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      28957 CACAACCACAACCACAAC 28940

RESULT 416
AC020112/c
LOCUS       AC020112               36708 bp    DNA     linear   HTG 03-JAN-2000
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC020112
VERSION     AC020112.1  GI:6664785
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 36708)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10212075 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .36708
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
ORIGIN

Alignment Scores:
Pred. No.:              7.52e+03        Length:         36708
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AC020112 (1-36708)
```

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      24050 CACAATCACAATCACAAC 24033

RESULT 417
AC149319/c
LOCUS       AC149319               36854 bp    DNA     linear   HTG 03-JUN-2004
DEFINITION  Phakopsora pachyrhizi clone JGIAFNA-101K17, WORKING DRAFT SEQUENCE,
            3 unordered pieces.
ACCESSION   AC149319
VERSION     AC149319.1  GI:48055852
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Phakopsora pachyrhizi
  ORGANISM  Phakopsora pachyrhizi
            Eukaryota; Fungi; Basidiomycota; Urediniomycetes;
            Urediniomycetidae; Uredinales; Phakopsoraceae; Phakopsora.
REFERENCE   1  (bases 1 to 36854)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Unpublished
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 36854)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUN-2004) Production Genomics Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA
            94598-1698, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 3634951
            Center clone name: PhakP-101K17
            -------
            Summary Statistics
            Consensus quality: 36415 bases at least Q40
            Consensus quality: 36526 bases at least Q30
            Consensus quality: 36578 bases at least Q20
            Estimated insert size: 40000; agarose-fp estimation
            Estimated insert size: 36654; sum-of-contigs estimation
            Quality coverage:  9.79 in Q20 bases; agarose-fp estimation
            Quality coverage:  10.68 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2126: contig of 2126 bp in length
            *     2127     2226: gap of unknown length
            *     2227    10691: contig of 8465 bp in length
            *    10692    10791: gap of unknown length
            *    10792    36854: contig of 26063 bp in length.
FEATURES             Location/Qualifiers
     source          1. .36854
                     /organism="Phakopsora pachyrhizi"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:170000"
                     /clone="JGIAFNA-101K17"
                     /clone_lib="JGI Fosmid library AFNA"
     gap             2127. .2226
                     /estimated_length=unknown
     gap             10692. .10791
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              7.54e+03        Length:         36854
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AC149319 (1-36854)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      33207 CACAACCACAACCACAAC 33190

RESULT 418
AC107074/c
LOCUS       AC107074               36933 bp    DNA     linear   PRI 08-JUN-2002
DEFINITION  Homo sapiens BAC clone RP11-424M21 from 4, complete sequence.
ACCESSION   AC107074
VERSION     AC107074.5  GI:21313973
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 36933)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
   PUBMED   9847074
REFERENCE   2  (bases 1 to 36933)
  AUTHORS   Kruchowski,S. and Kozlowicz,A.
  TITLE     The sequence of Homo sapiens BAC clone RP11-424M21
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 36933)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JAN-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 36933)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-APR-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 36933)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   6  (bases 1 to 36933)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Jun 2, 2002 this sequence version replaced gi:20177786.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0424M21
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
```

as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at http://www.chori.org
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is RP11-503L23, 2000 bp overlap. Actual start of this clone is at base position 1 of RP11-424M21; actual end is at base position 85936 of RP11-73G16.

There is a single plasmid subclone region from 14351 to 14470. Polymorphisms have been identified between AC107074 and AC044853. Data from AC044853 was used to finish this clone, AC107074.

FEATURES Location/Qualifiers

source 1. .36933
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-424M21"
/clone_lib="RPCI-11"

repeat_region 710. .772
/rpt_family="(TTCA)n"

repeat_region 822. .871
/rpt_family="L2"

repeat_region 1500. .1578
/rpt_family="L2"

repeat_region 4179. .4417
/rpt_family="L1"

repeat_region 4419. .4441
/rpt_family="AT_rich"

repeat_region 6023. .6305
/rpt_family="Alu"

repeat_region 6447. .6673
/rpt_family="L2"

repeat_region 6704. .7108
/rpt_family="L2"

repeat_region 7310. .7600
/rpt_family="Alu"

repeat_region 8681. .9453
/rpt_family="L2"

repeat_region 10334. .10727
/rpt_family="MaLR"

repeat_region 12487. .12534
/rpt_family="L2"

repeat_region 12952. .13462
/rpt_family="MaLR"

repeat_region 13540. .13593
/rpt_family="MIR"

repeat_region 14370. .14391
/rpt_family="(TCCCC)n"

repeat_region 14431. .14463
/rpt_family="AT_rich"

repeat_region 14513. .15647
/rpt_family="MER2_type"

repeat_region 15648. .15805
/rpt_family="MaLR"

repeat_region 15850. .16051
/rpt_family="L2"

repeat_region 16104. .16470
/rpt_family="MaLR"

repeat_region 16861. .16900
/rpt_family="MER1_type"

repeat_region 16901. .17426
/rpt_family="ERVL"

repeat_region 17427. .17500
/rpt_family="MER1_type"

repeat_region 17606. .17632
/rpt_family="(TATG)n"

repeat_region 17637. .17712
/rpt_family="L2"

repeat_region 17795. .18199
/rpt_family="MaLR"

repeat_region 18200. .18495
/rpt_family="Alu"

repeat_region 18496. .18559
/rpt_family="MaLR"

repeat_region 20372. .20514
/rpt_family="MER1_type"

repeat_region 21770. .21872
/rpt_family="MIR"

repeat_region 22524. .22624
/rpt_family="L2"

repeat_region 22721. .22827
/rpt_family="L2"

repeat_region 22861. .23360
/rpt_family="MaLR"

repeat_region 23438. .23545
/rpt_family="MER1_type"

repeat_region 23671. .23714
/rpt_family="L2"

repeat_region 24283. .24332
/rpt_family="(TCCC)n"

repeat_region 24332. .24356
/rpt_family="(T)n"

repeat_region 24407. .24585
/rpt_family="MER1_type"

repeat_region 24662. .24968
/rpt_family="Alu"

repeat_region 25264. .25637
/rpt_family="ERV1"

repeat_region 25638. .25853
/rpt_family="ERV1"

repeat_region 25855. .26517
/rpt_family="ERV1"

repeat_region 26520. .27012
/rpt_family="ERV1"

repeat_region 30208. .30610
/rpt_family="ERV1"

repeat_region 30617. .30711
/rpt_family="AcHobo"

repeat_region 31522. .31552
/rpt_family="ERV1"

repeat_region 31553. .31608
/rpt_family="(TCTA)n"

repeat_region 31609. .31836
/rpt_family="ERV1"

repeat_region 31842. .32348
/rpt_family="ERV1"

repeat_region 32408. .32727
/rpt_family="MER1_type"

repeat_region 33923. .34024
/rpt_family="L2"

Alignment Scores:

Pred. No.:	7.56e+03	Length:	36933
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0

```
DB:                    8             Gaps:          0

US-09-858-332G-28 (1-6) x AC107074 (1-36933)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      17796 CATAACCATAACCATAAT 17779

RESULT 419
AF040647/c
LOCUS       AF040647               38094 bp    DNA     linear   INV 27-MAY-2005
DEFINITION  Caenorhabditis elegans cosmid F54D12, complete sequence.
ACCESSION   AF040647
VERSION     AF040647.1  GI:2746827
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 38094)
  AUTHORS   .
  CONSRTM   C. elegans Sequencing Consortium
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology
  JOURNAL   Science 282 (5396), 2012-2018 (1998)
   PUBMED   9851916
REFERENCE   2  (bases 1 to 38094)
  AUTHORS   Maggi,L. and Goela,D.
  TITLE     The sequence of C. elegans cosmid F54D12
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 38094)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-DEC-1997) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   4  (bases 1 to 38094)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAY-2002) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   5  (bases 1 to 38094)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2002) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   6  (bases 1 to 38094)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-2002) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   7  (bases 1 to 38094)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2003) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   8  (bases 1 to 38094)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-2003) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   9  (bases 1 to 38094)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JUN-2004) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   10 (bases 1 to 38094)
  AUTHORS   .
  CONSRTM   WormBase Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (21-SEP-2004) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   11 (bases 1 to 38094)
  AUTHORS   .
  CONSRTM   WormBase Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAY-2005) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
COMMENT     Submitted by:
                  Genome Sequencing Center
                  Department of Genetics, Washington University
                  St. Louis , MO  63110, USA, and
                  Sanger Centre, Hinxton Hall
                  Cambridge CB10 IRQ, England
                  email: submissions@watson.wustl.edu and jes@sanger.ac.uk

            NOTICE:  This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate chemistry
            or covered by high quality data (i.e., phred quality >= 30); an
            attempt was made to resolve all sequencing problems, such as
            compressions and repeats; all regions were covered by sequence from
            more than one m13 subclone.

            For a graphical representation of this clone sequence and its
            analysis see:
            http://www.wormbase.org/db/seq/sequence?name=F54D12;class=Sequence

                        NEIGHBORING CLONE INFORMATION

            The 5' clone is Y51H7D, 200 bp overlap; the 3' clone is Y51H7C, 200
            bp overlap. Actual start of this clone is at base position 1 of
            F54D12; actual end is at 38094 of F54D12.

            NOTES:

            Coding seqences below are the result of integration and manual
            review of the following data : computer analysis using the program
            Genefinder (P. Green and L. Hillier, personal communication), the
            large scale EST projects of Yuji Kohara
            (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C.
            elegans ORFeome cloning project (http://worfdb.dfci.harvard.edu/),
            similarity to other proteins from BlastX analyses
            (http://blast.wustl.edu/), sequence conservation with C. briggsae
            using Jim Kent's WABA alignment program (Genome Research
            10:1115-1125, 2000), individual C. elegans GenBank submissions,
            and personal communications with C. elegans researchers. tRNAs
            are predicted using the program tRNAscan-SE (Lowe, T.M. and
            Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).
FEATURES                Location/Qualifiers
     source             1. .38094
                        /organism="Caenorhabditis elegans"
                        /mol_type="genomic DNA"
                        /strain="Bristol N2"
                        /db_xref="taxon:6239"
                        /chromosome="II"
                        /clone="F54D12"
     tRNA               complement(3047. .3120)
                        /note="F54D12.t1"
                        /pseudo
                        /db_xref="WormBase:F54D12.t1"
     tRNA               4592. .4664
                        /note="F54D12.t2"
                        /pseudo
                        /db_xref="WormBase:F54D12.t2"
```

```
     gene            complement(4741. .4950)
                     /locus_tag="F54D12.4"
     CDS             complement(join(4741. .4814,4887. .4950))
                     /locus_tag="F54D12.4"
                     /standard_name="F54D12.4"
                     /note="contains similarity to Lymantria dispar multicapsid
                     nuclear polyhedrosis virus LdOrf-117 peptide.; TR:Q9YML0"
                     /codon_start=1
                     /product="Hypothetical protein F54D12.4"
                     /protein_id="AAB94994.1"
                     /db_xref="GI:2746833"
                     /db_xref="WormBase:F54D12.4"
                     /translation="MSTNSFVSSFFIAIPQPVFCQRANFLCYILNSDMMNELFELAQL
                     C"
     gene            complement(5697. .9157)
                     /locus_tag="F54D12.5"
     CDS             complement(join(5697. .5796,6467. .6750,8172. .8637,
                     9108. .9157))
                     /locus_tag="F54D12.5"
                     /standard_name="F54D12.5"
                     /note="contains similarity to Streptococcus pneumoniae
                     Cell wall surface anchor family protein.; TR:Q97P71; coded
                     for by the following C. elegans cDNAs: OSTF113D12_1,
                     OSTF123C3_1, OSTR113D12_1, yk1108a06.5, OSTR123C3_1"
                     /codon_start=1
                     /product="Hypothetical protein F54D12.5"
                     /protein_id="AAB94995.3"
                     /db_xref="GI:49035118"
                     /db_xref="WormBase:F54D12.5"
                     /translation="MGGGGATRGSKLLGNIVTERTNPKASGISNGILPPYINISPKEV
                     PSSSRHHYYAATSSSSMTASAASEVSAALSAMNRDISSGSGSSTAASATATTSSKLPR
                     CTPPPVSNRQLKRYAMVPGTSMSKRRISREDAKRAFLSSDPRVGHVIDDPRYAAVNVE
                     ADENYANLRDIQTDKIMRGIDVETLTNLHTTTEPSPPLPRHHQIQASTTAGATSSSLS
                     AAAGTRGTLKPSNGLMTARAATASPLPRLNYAQIVPVHDDGVVPDRESRCSSVGRFFD
                     INYAQMDADRTQALKGTLRQKKE"
     gene            complement(10340. .11636)
                     /locus_tag="F54D12.6"
     CDS             complement(join(10340. .10576,11055. .11248,11480. .11636))
                     /locus_tag="F54D12.6"
                     /standard_name="F54D12.6"
                     /note="contains similarity to Pfam domain PF02174 (PTB
                     domain (IRS-1 type))"
                     /codon_start=1
                     /product="Hypothetical protein F54D12.6"
                     /protein_id="AAB94991.1"
                     /db_xref="GI:2746830"
                     /db_xref="WormBase:F54D12.6"
                     /translation="MRAIISVAKSILGNRRGDDESLYTGAGPAGSGASHSFRVYIHGN
                     RKKNELVHAWLRVTRTTITLEISKKECLVWPLPLIRRYGYTSAGIFFFESGRRCESGE
                     GMFTFQSKKADEIFQLIQSLIEEFANTSMEKQNMENSFRYQRQQSVPISSRRMSNTSN
                     TSSVFGTADYKLQVSKLKNGSARLQTTNFDLADPN"
     gene            complement(14279. .15010)
                     /locus_tag="F54D12.10"
     CDS             complement(join(14279. .14452,14726. .15010))
                     /locus_tag="F54D12.10"
                     /standard_name="F54D12.10"
                     /codon_start=1
                     /product="Hypothetical protein F54D12.10"
                     /protein_id="AAY55829.1"
                     /db_xref="GI:66773609"
                     /db_xref="WormBase:F54D12.10"
                     /translation="MSSQTFLILFLFPITILSVFVSVRGSPVNSTSYVSKTFLNLTWK
                     SCITRCVIVADCILVHSNSLNRCYLYAVGDIIQVRNDRDGYSIMNETVAFRMRNSPYK
                     CSNRSSDMLFGVINLYNKDNITSYEITMSPTNEYYEIKYGRSKLLEISNV"

Alignment Scores:
Pred. No.:              7.79e+03        Length:         38094
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-858-332G-28 (1-6) x AF040647 (1-38094)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        269 CATAATCATAATCATAAT 252

RESULT 420
CEF35H8
LOCUS       CEF35H8                38383 bp    DNA     linear   INV 09-AUG-2005
DEFINITION  Caenorhabditis elegans Cosmid F35H8, complete sequence.
ACCESSION   Z36752
VERSION     Z36752.1  GI:1066957
KEYWORDS    HTG; Arg-tRNA; Beta-1,6-N-acetylglucosaminyltransferase; HUD;
            Integrin alpha-V; Paraneoplastic encephalomyelitis antigen; serine
            threonine kinase; transfer RNA; UDP-glucuronosyltransferase; Wee1;
            Zinc finger protein.
SOURCE      Caenorhabditis elegans
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 38383)
  AUTHORS   .
  CONSRTM   C. elegans Sequencing Consortium
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology
  JOURNAL   Science 282 (5396), 2012-2018 (1998)
   PUBMED   9851916
REFERENCE   2  (bases 1 to 38383)
  AUTHORS   Berks,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-AUG-1994) Nematode Sequencing Project, Sanger
            Institute, Hinxton, Cambridge CB10 1SA, England and Department of
            Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
            worm@sanger.ac.uk
COMMENT     On Nov 21, 1995 this sequence version replaced gi:531770.
            Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT:  This sequence is NOT necessarily the entire insert of
            the specified clone.  It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
            For a graphical representation of this sequence and its analysis
            see:- http://www.wormbase.org/perl/ace/elegans/seq/sequence?
            name=F35H8;class=Sequence
            IMPORTANT: This sequence is not the entire insert of clone F35H8.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone F35H8 is at 1 in this sequence. The true
            right end of clone F35H8 is at 96 in
            sequence Z47075.
            The true left end of clone E02H1 is at 14958 in this sequence. The
            true right end of clone AH6 is at 3246 in this sequence. The start
            of this sequence (1. .116) overlaps with the end of sequence Z48009.
            The end of this sequence (38288. .38383) overlaps with the start of
            sequence Z47075.
FEATURES             Location/Qualifiers
     source          1. .38383
                     /organism="Caenorhabditis elegans"
                     /mol_type="genomic DNA"
                     /strain="Bristol N2"
                     /db_xref="taxon:6239"
                     /chromosome="II"
                     /clone="F35H8"
     gene            complement(join(2409. .2483,2528. .2635,2683. .2753,
                     2800. .2998))
                     /locus_tag="F35H8.1"
```

```
CDS             complement(join(2409. .2483,2528. .2635,2683. .2753,
                2800. .2998))
                /locus_tag="F35H8.1"
                /standard_name="F35H8.1"
                /note="contains similarity to Plasmodium falciparum
                Hypothetical protein.; TR:Q8ILJ6"
                /codon_start=1
                /product="Hypothetical protein F35H8.1"
                /protein_id="CAA85323.2"
                /db_xref="GI:37619785"
                /db_xref="UniProt/TrEMBL:Q20080"
                /translation="MRYLSNLLCLTLTFIVCFGAERDADEEAYVLMNLLKQSIDMDQT
                EALPGFVSRDFTYDTCKGIWNRDKYMEYLKVFSSRPDGGMQPMSTDWKVTNAHFNDNT
                LYFNIHNYNQHLAIYPDLEFEAKKIGKDERYILTRGKISKCGTLEKEY"
gene            join(5262. .5365,5418. .5549,5598. .5694,5741. .5989,
                6035. .6114,6392. .6450,6499. .6721,6968. .7139)
                /locus_tag="F35H8.2"
CDS             join(5262. .5365,5418. .5549,5598. .5694,5741. .5989,
                6035. .6114,6392. .6450,6499. .6721,6968. .7139)
                /locus_tag="F35H8.2"
                /standard_name="F35H8.2"
                /note="contains similarity to Pfam domain PF02485
                (Core-2/I-Branching enzyme)"
                /codon_start=1
                /product="Hypothetical protein F35H8.2"
                /protein_id="CAA85324.1"
                /db_xref="GI:3876750"
                /db_xref="GOA:Q20081"
                /db_xref="InterPro:IPR003406"
                /db_xref="UniProt/TrEMBL:Q20081"
                /translation="MANHDFCTKTMWALSVPLGVLLILFIALTAYATYSFHIFVPIAP
                NNEIDCAIMFNKSSEDYLTYIENRTLLETTHNDRNCFQIRRRRYLPADLPKSMSGFHN
                MFFIRVVSKDYDFVEEVLAMMHSPIHFFCFVLDINSEPLFKERMFRLGDCMVNVLVPR
                ELFNTSTAHGTLNAHRRCLREIDEFDWKHAVITAEHDIPLHSTKFLSRRSRRLRDMVE
                MNGIHFKEETLANANPNDTSSQLFQVSRNWLQHLCSVATISHSQHKVLYDYLLKTPDF
                DLPNDQHREISVVEQKCDTMKYDVDGSCLYAMEDFSAIRNTTRLFIRADPHYDFGFVQ
                CVHEVVFERTFFDPISNKNANNYLYGSRVNLIDDYRN"
gene            join(7491. .7560,7619. .7703,7749. .7885,7948. .8228,
                8272. .8520,8580. .8860,8929. .9094)
                /locus_tag="F35H8.3"
CDS             join(7491. .7560,7619. .7703,7749. .7885,7948. .8228,
                8272. .8520,8580. .8860,8929. .9094)
                /locus_tag="F35H8.3"
                /standard_name="F35H8.3"
                /note="contains similarity to Pfam domain PF00096 (Zinc
                finger, C2H2 type)"
                /codon_start=1
                /product="Hypothetical protein F35H8.3"
                /protein_id="CAA85325.1"
                /db_xref="GI:3876751"
                /db_xref="GOA:Q20082"
                /db_xref="InterPro:IPR007087"
                /db_xref="UniProt/TrEMBL:Q20082"
                /translation="MEEMMMNDPSAMVIYEEEVTTAPNLPCSLIQQRSWDEEKPIGYE
                LTNTKCYKTPNGVQKTATIQREQLSTSKDFGEQQIVEMDGEFSIEGTDMHAIPCTSSS
                MQPSTSSNPSSGEHQPVPLRRMAIKIGQRVLRFKVISAEEAPEAPLDTQDSWINDPKP
                VTTPKALAGLYRCTNCKTYFGNKEVYQRHIQEVHGDARPFRCFNCGMRFANKTSMTHH
                LKDHSLLKPMFSCDYCPRIFSKLESKTRHHKMHFTRSTCQTCMRFFTTEDALRHHQST
                AHPATFDSGPPPEDLLPNGKSARYSCSYCNLRFHFKKDMLVHERIHTGEKPYSCGYCM
                KSFAQSQALTAHIRTHTKELPYGCGKCDKRFRDNSCLRKHELAAHTDEPIVRPISVAY
                SNQVQKQMQRQRENRRKQELLIAERHPYRI"
gene            complement(join(9280. .9411,9465. .9572,9621. .9686))
                /locus_tag="F35H8.4"
CDS             complement(join(9280. .9411,9465. .9572,9621. .9686))
                /locus_tag="F35H8.4"
                /standard_name="F35H8.4"
                /note="contains similarity to Streptococcus pyogenes
                Hypothetical protein.; TR:Q8P0S2"
                /codon_start=1
                /product="Hypothetical protein F35H8.4"
                /protein_id="CAA85326.2"
                /db_xref="GI:37619784"
                /db_xref="UniProt/TrEMBL:Q20083"
                /translation="MATTNVTMMSVKSLTTTESDIEDGPNPTPPYVYILIEVGYIVLF
                VIVTYLMRKYNFMKTDDDAVTKHHRRHFEMMKRQQEDLITRQKSNENFLVANNEGKA"
gene            15988. .16060
                /locus_tag="F35H8.t1"
tRNA            15988. .16060
                /locus_tag="F35H8.t1"
                /standard_name="F35H8.t1"
                /note="preliminary prediction
                Predicted tRNA, Cove score: 73.16; Hypothetical RNA
                transcript F35H8.t1"
gene            join(26055. .26114,28035. .28212,28939. .29042,29546. .29699,
                29808. .30099,30194. .30617,30666. .30824)
                /gene="exc-7"
                /locus_tag="F35H8.5"
CDS             join(26055. .26114,28035. .28212,28939. .29042,29546. .29699,
                29808. .30099,30194. .30617,30666. .30824)
                /gene="exc-7"
                /locus_tag="F35H8.5"
                /standard_name="F35H8.5"
                /note="C. elegans EXC-7 protein; contains similarity to
                Pfam domain PF00076 (RNA recognition motif. (aka RRM, RBD,
                or RNP domain))"
                /codon_start=1
                /product="Hypothetical protein F35H8.5"
                /protein_id="CAA85327.1"
                /db_xref="GI:3876753"
                /db_xref="GOA:Q20084"
                /db_xref="HSSP:1FXL"
                /db_xref="InterPro:IPR000504"
                /db_xref="InterPro:IPR002343"
                /db_xref="InterPro:IPR006548"
                /db_xref="InterPro:IPR012677"
                /db_xref="UniProt/TrEMBL:Q20084"
                /translation="MTMEALLTPNNNNVPTSPVQTTYTSTQRYCSAPNVDIGESKTNL
                IINYLPQGMTQEEVRSLFTSIGEIESCKLVRDKVTGQSLGYGFVNYVREEDALRAVSS
                FNGLRLQNKTIKVSYARPSNDQIKGSNLYVSGIPKSMTLHELESIFRPFGQIITSRIL
                SDNVTGLSKGVGFVRFDKKDEADVAIKTLNGSIPSGCSEQITVKFANNPASNNPKGLL
                SDLEAVQQAATTLVPLSTILGAPTLRATAGGIGPMHHAPITSKYRYSPMGAITAVSQP
                TATLPADYLTTSALLQMSQLNALAGLNPFATATAVPDFTASLIAHQQQQHAVAQQHAA
                QTASPPATNGQVAGLAAHAQLSALSASVAATLPPSDTAGYCLFVYNLSSDTDDTLLWQ
                LFSQFGAIVNVKILRDLTQQCKGYAFVSMSNYTEAYNAMLSLNGTNLAGKTLQVVFKS
                STPYRA"
gene            complement(join(31711. .32001,32054. .32416,32460. .32962,
                33045. .33294))
                /gene="wee-1.1"
                /locus_tag="F35H8.7"
CDS             complement(join(31711. .32001,32054. .32416,32460. .32962,
                33045. .33294))
                /gene="wee-1.1"
                /locus_tag="F35H8.7"
                /standard_name="F35H8.7"
                /note="C. elegans WEE-1.1 protein; contains similarity to
                Pfam domain PF00069 (Eukaryotic protein kinase domain)"
                /codon_start=1
                /product="Hypothetical protein F35H8.7"
                /protein_id="CAA85329.1"
                /db_xref="GI:3876755"
                /db_xref="GOA:Q20085"
                /db_xref="InterPro:IPR000719"
                /db_xref="InterPro:IPR008271"
                /db_xref="UniProt/Swiss-Prot:Q20085"
                /translation="MNSVLSNIEKLDTYILRRNEYVAESKDEPNKLNTSRKLEVTTKK
```

Alignment Scores:

Pred. No.:	7.85e+03	Length:	38383
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-28 (1-6) x CEF35H8 (1-38383)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      18475 CATAATCATAATCATAAT 18492

RESULT 421
AC152111
LOCUS       AC152111               38449 bp    DNA     linear   PLN 01-DEC-2004
DEFINITION  Pichia stipitis clone JGIAHYZ-17L1, complete sequence.
ACCESSION   AC152111
VERSION     AC152111.1  GI:54654145
KEYWORDS    HTG.
SOURCE      Pichia stipitis
  ORGANISM  Pichia stipitis
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Pichia.
REFERENCE   1  (bases 1 to 38449)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 38449)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-OCT-2004) Production Genomics Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA
            94598-1698, USA
REFERENCE   3  (bases 1 to 38449)
  AUTHORS   Stanford Human Genome Center.
  CONSRTM   DOE Joint Genome Institute
  TITLE     Direct Submission
  JOURNAL   Submitted (01-DEC-2004) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.9% of Sequence;
            Estimated Total Number of Errors is 0.
FEATURES             Location/Qualifiers
     source          1. .38449
                     /organism="Pichia stipitis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4924"
                     /clone="JGIAHYZ-17L1"
ORIGIN

Alignment Scores:
Pred. No.:              7.86e+03        Length:         38449
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x AC152111 (1-38449)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        810 CACAACCATAACCATAAC 827

RESULT 422
AR408742
LOCUS       AR408742               39003 bp    DNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 21 from patent US 6632636.
ACCESSION   AR408742
VERSION     AR408742.1  GI:40159137
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 39003)
  AUTHORS   Lagace,R.E., Patterson,C. and Berg,K.L.
  TITLE     Nucleic acids encoding 3-ketoacyl-ACP reductase from Moraxella
            catarrahalis
  JOURNAL   Patent: US 6632636-A 21 14-OCT-2003;
            Elitra Pharmaceuticals Inc.; San Diego, CA
FEATURES             Location/Qualifiers
     source          1. .39003
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

Alignment Scores:
Pred. No.:              7.97e+03        Length:         39003
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x AR408742 (1-39003)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       3052 CACAATCATAATCATAAT 3069

RESULT 423
AX067446
LOCUS       AX067446               39003 bp    DNA     linear   PAT 24-JAN-2001
DEFINITION  Sequence 21 from Patent WO0078968.
ACCESSION   AX067446
VERSION     AX067446.1  GI:12545066
KEYWORDS    .
SOURCE      Moraxella catarrhalis
  ORGANISM  Moraxella catarrhalis
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Moraxellaceae; Moraxella.
REFERENCE   1
  AUTHORS   Lagace,R.E., Patterson,C. and Berg,K.L.
  TITLE     Nucleotide sequences of moraxella catarrhalis genome
  JOURNAL   Patent: WO 0078968-A 21 28-DEC-2000;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .39003
                     /organism="Moraxella catarrhalis"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:480"
ORIGIN

Alignment Scores:
Pred. No.:              7.97e+03        Length:         39003
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x AX067446 (1-39003)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       3052 CACAATCATAATCATAAT 3069

RESULT 424
AC156596/c
LOCUS       AC156596               39074 bp    DNA     linear   HTG 31-JAN-2005
DEFINITION  Laccaria bicolor clone JGIAHZW-15X, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION   AC156596
VERSION     AC156596.1  GI:58338252
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Laccaria bicolor
  ORGANISM  Laccaria bicolor
            Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
            Agaricales; Tricholomataceae; Laccaria.
REFERENCE   1  (bases 1 to 39074)
  AUTHORS   DOE Joint Genome Institute.
```

```
  TITLE     Unpublished
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 39074)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2005) Production Genomics Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA
            94598-1698, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 3640135
            Center clone name: JGI-AHZW_15X
            -------
            Summary Statistics
            Consensus quality: 38787 bases at least Q40
            Consensus quality: 38932 bases at least Q30
            Consensus quality: 38960 bases at least Q20
            Estimated insert size: 40000; agarose-fp estimation
            Estimated insert size: 38974; sum-of-contigs estimation
            Quality coverage:  7.26 in Q20 bases; agarose-fp estimation
            Quality coverage:  7.45 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    11646: contig of 11646 bp in length
            *   11647    11746: gap of unknown length
            *   11747    39074: contig of 27328 bp in length.
FEATURES            Location/Qualifiers
     source         1. .39074
                    /organism="Laccaria bicolor"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:29883"
                    /clone="JGIAHZW-15X"
                    /clone_lib="JGI Fosmid library AHZW"
     gap            11647. .11746
                    /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              7.98e+03        Length:         39074
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AC156596 (1-39074)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      12612 CACAACCACAACCATAAC 12595

RESULT 425
AC165067/c
LOCUS       AC165067               39156 bp    DNA     linear   PLN 26-JUL-2005
DEFINITION  Phakopsora pachyrhizi clone JGIAFNA-421A12, complete sequence.
ACCESSION   AC165067
VERSION     AC165067.2  GI:71143210
KEYWORDS    HTG.
SOURCE      Phakopsora pachyrhizi
  ORGANISM  Phakopsora pachyrhizi
            Eukaryota; Fungi; Basidiomycota; Urediniomycetes;
            Urediniomycetidae; Uredinales; Phakopsoraceae; Phakopsora.
REFERENCE   1  (bases 1 to 39156)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 39156)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-2005) Production Genomics Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA
            94598-1698, USA
REFERENCE   3  (bases 1 to 39156)
  AUTHORS   Stanford Human Genome Center.
  CONSRTM   DOE Joint Genome Institute
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2005) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jul 26, 2005 this sequence version replaced gi:68533351.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 100% of Sequence;
            Estimated Total Number of Errors is 0.
FEATURES            Location/Qualifiers
     source         1. .39156
                    /organism="Phakopsora pachyrhizi"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:170000"
                    /clone="JGIAFNA-421A12"
ORIGIN

Alignment Scores:
Pred. No.:              8e+03           Length:         39156
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x AC165067 (1-39156)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      18986 CATAACCACAACCACAAC 18969

RESULT 426
AC020271
LOCUS       AC020271               39261 bp    DNA     linear   HTG 03-JAN-2000
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC020271
VERSION     AC020271.1  GI:6664626
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 39261)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10212625 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES            Location/Qualifiers
     source         1. .39261
                    /organism="Drosophila melanogaster"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:7227"
ORIGIN
```

```
Alignment Scores:
Pred. No.:              8.02e+03        Length:         39261
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AC020271 (1-39261)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       6055 CATAATCATAATCATAAT 6072

RESULT 427
AC116925/c
LOCUS       AC116925               39575 bp    DNA     linear   INV 12-MAR-2003
DEFINITION  Dictyostelium discoideum chromosome 2 map 4189423-4228998 strain
            AX4, complete sequence.
ACCESSION   AC116925
VERSION     AC116925.2  GI:28829732
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 39575)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 39575)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 39575)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 39575)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On Mar 4, 2003 this sequence version replaced gi:19909383.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES             Location/Qualifiers
     source          1. .39575
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /chromosome="2"
                     /map="4189423-4228998"
     CDS             join(323. .335,460. .641)
                     /note="GeneID exon scores (in order of location ranges):
                     0.84, 3.48 - GSCJ_ID dd_03415"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO52236.1"
                     /db_xref="GI:28829733"
                     /translation="MTIIKSISIIADKSISKTNFIKINNENDSIVIMYTNNQKISKSR
                     GQILDSHIVVNPLIDFNLIF"
     CDS             complement(join(718. .917,1011. .1023))
                     /note="GeneID exon scores (in order of location ranges):
                     2.00, -0.18 - GSCJ_ID dd_03416"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO52237.1"
                     /db_xref="GI:28829734"
                     /translation="MTIINSISSLGKITNKNKSQNNLNSINNSINPPNNIQGSNEKTG
                     LVSGLLELVGGVVLILGLVVGATLGL"
     CDS             join(2320. .2803,2942. .3711)
                     /note="GeneID exon scores (in order of location ranges):
                     22.58, 56.62 - GSCJ_ID dd_00034"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO52238.1"
                     /db_xref="GI:28829735"
                     /translation="MELLKIIIKFFLKLKVQILVLGLIAWAVAFSIDRTLVEDYRNKS
                     DDIFTNENNFKMVRTNRESAAFIILSIFETGLFGLGIGFFVSLGFRVFKIKNKMLRKR
                     MIWCWAAISYFCVTWWPHSTSHFLILTTGRSPIDNYITMETTFHWTLTFCSSTLAYFQ
                     YDLIQLMRDVAIMRKRMVASGKEEVDPLKQEWYKNYHYHSIVIAIIVLAIAIPVAFHF
                     DPLTDKLERYQRFFYVTFHIVNSAVAAVGFAFAYAVTRTILLLPPSPGRKVSIVSSLA
                     ITFLFIIQYIHPMVHVKTKPSMKNTIAIDFGFHVPIVLATGILAFYQFRMMELATDDK
                     SSLSIAVRMRSDSKSASKGSSLKSADNASQSQSSNSFSTSIQIDHLKTKTPPIDESIN
                     NSSIEMKDLETNSKDNNISICIEKE"
     CDS             join(4657. .4669,4749. .4960)
                     /note="GeneID exon scores (in order of location ranges):
                     1.95, 8.09 - GSCJ_ID dd_00035"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO52239.1"
                     /db_xref="GI:28829736"
                     /translation="MTISGSFSSFGTTNSLNVKNNQNHFLSTSPNLNGNSILSSSFGG
                     SSSFQNSNYLIASVNSTVITESFKLLDNPW"
     CDS             complement(5219. .6355)
                     /note="GeneID exon scores (in order of location ranges):
                     77.28 - GSCJ_ID dd_00036"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO52240.1"
                     /db_xref="GI:28829737"
                     /translation="MVTTNNNNKNNKIKTINNDKLKSKSTTKIKSKSKCLNKFIIYT
                     SISILIISIIISIGFIGFTSHSLLYFPWLRTQHKPNIGDPKTKFGIDFQDIEFPSFDG
                     ENNNNNNSSMKIRGWWIPASKENCKQLCVISIHGSGRDRYEWINHTKLFHDEGISILN
                     YDSIDGVGKSDSANRGVGYSYREHKDVRSAIRLLKTTEPYNKICKKLLLTGMSLGGGS
                     VIIAAAKDRHLIDGVIAESPYALASSAWRHNIFRFLGAGISSFFPFIKSPYDQLLPKN
                     PPEIIVDTILKLTEYRIISTTINVSKDDTPLNYVDKISPKPILFIHATHDPVVPYSQS
                     VLLYEKANHPKELWTVHNKTHQTAFHSTSIDQYKDKLRNFIKQL"
     CDS             complement(join(6817. .7338,7512. .7673))
                     /note="GeneID exon scores (in order of location ranges):
                     66.55, 14.67 - GSCJ_ID dd_00037"
                     /codon_start=1
                     /product="similar to Rhizobium loti (Mesorhizobium loti).
                     Glutathione S-transferase"
                     /protein_id="AAO52241.1"
                     /db_xref="GI:28829738"
                     /translation="MSHLTPKPNVLNVYGAGTPNFVKVQILLEEKNIPHVYHKIDFSK
                     NEQFDPEFLKFSPNNKIPAIVDNTVEPNGISVFESAAILMYLSEKHQFTDLYADFKTD
                     PALHAKHLEFLFFQMSGVGPMTGQLGWFLRQPEPKNTQAIERYSEEVKRLYRVLDKIL
                     STNTWVAGEQYSIADISLLGWALYFLKGYYDDHITREQFPNVYRWLDLAAKRPAIIKV
                     LETIQPKQK"
     CDS             join(8976. .9290,9396. .11147)
                     /note="GeneID exon scores (in order of location ranges):
                     33.56, 170.11 - GSCJ_ID dd_00038"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO52242.1"
                     /db_xref="GI:28829739"
                     /translation="MGRSPEETQPLVGGGGGGKSDPYSFMYPGSGSISTLNINGSSNS
                     YPLDNVLQHKGSDYIGENNDVLVGTEAKMVLTGEVENKKVLDGSDEHKIGQLFSTAIC
                     GNDITSSIFYTSSLCTAAAGKYAPVSLILVCIVLYLFRKVYGEVGSALPLNGGAYNVL
```

```
                LNTTTKQVAALAAALTMISYMATAVVSASSAMSYLKVLWEGCDVYWATIILLGGFAAL
                NLIGISESAVVAFGIFALHMLTLAMFVIGGFVLFFIQRESTFQIIRDNWYNPEGNCPV
                GTPTPYYFGKNFAEAIYFGFGAAMLGVTGFETSSNFIEQQEPGVFPKTLRNMWFIVSI
                FNPLIGFLNLCFIPACEIASDTSGGILAKVAGKAFGAWFEKWVGIDAVLVLSGSVITS
                YVGVIGLVRRMALDRVLPQFLIQTNPCRKTTHWIIVLFFVICSSLYMLVDGDTTTLAG
                VYTCAFLGVMSLFAIGNMLLKYKRSSLRRDIKSPWIGVLIALFSVILGLISNIIKDDS
                IVYYFGIYFGVTIFIIMLMFIRARILKLFLFFTQIILSGAGRDRVARWIGDVIKKINS
                QPVVFFASKDDPAYLNKAILYIRDNEQTNWVRVVHCCNEEESTDSFSSTMEFLDKCYP
                KIRLDLIRVNAEFNASTVQKISQELKIDLNFCFVSCIKDRDDGDPSMIPVTVGDLGGV
                RMITSA"
     CDS        complement(join(11476. .11696,11800. .11809))
                /note="GeneID exon scores (in order of location ranges):
                20.11, -0.81 - GSCJ_ID dd_00039"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAO52243.1"
                /db_xref="GI:28829740"
                /translation="MLFESISTLSNLKSASKSSMIASTGSTSSKSSNSIQCYTGCVGG
                GIFPIRQTSINIDINISISRGCGSCGGCPFKY"
     CDS        complement(join(13183. .13433,13545. .13557))
                /note="GeneID exon scores (in order of location ranges):
                13.58, 3.28 - GSCJ_ID dd_03365"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAO52244.1"
                /db_xref="GI:28829741"
                /translation="MAILASICSIGNIKTNSSKLIINNISSQSSNMNYNIIQCNGPMS
                NSSGCGNSTGPIVGGVIINRGAVSGVLCIVGGVVNGLVGGGCN"
     CDS        complement(14739. .15287)
                /note="GeneID exon scores (in order of location ranges):
                31.57 - GSCJ_ID dd_00780"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAO52245.1"
                /db_xref="GI:28829742"
                /translation="MKLSNSILILFVFLNISIVKAEYKSFLIQMDGCNNIMETNIGQC
                TTNKCNDTTSFKITHVSGNLYELYLYKNKCETGINYSFINCSTIPIENQNGQNGQNSI
                LPQINLNGIEANINDGKISLSKLNSKIYCFNEMSHSSHDDDFLQLNNFNHHETDLSSN
                SIKHSKYNFITLFSYLLVLLII"
     CDS        complement(15799. .19152)
                /note="GeneID exon scores (in order of location ranges):
                241.32 - GSCJ_ID dd_00781"
                /codon_start=1
                /product="similar to Dictyostelium discoideum (Slime
                mold). HelE (Fragment)"
                /protein_id="AAO52246.1"
                /db_xref="GI:28829743"
                /translation="MKILIKRIEILKKRWRMFWKSYFIRQMIVFLQVMILIYLINKAV
                TKFQMLDNKTTIQEFHNDECKINPTTGKINNSDCLKTGNQFKRVLFFYNDGLARDLSD
                ELISMFKQSSNTFRIQNEGFPASYAVFSSYVTGTPPTNFMGNPILSDNLIYQLTSNGI
                PMGYFGTRMPAYDVLEEGKYFEKDQVNIEYLKEGSQLYENLFKCAKNSPEKCSQQFLD
                QKKEKGLSIFFSGDIVDERNHHTGDKYDKTTISIINEWFGNMKYIKQWVDENPEYLLV
                IFSDHGGKLTSETAAHGANNGGNEAFMVIYNPTIDPLPSNLQDKFIDQVDVCSTIWQH
                FSGSGVSLPSENIGKISPTTTDPERIYYTYKLNANQLNRFGKRWNYKLRTSDYNQAVE
                TMNKDINHSIELFIHYINSLKIPFLNLKKFPTTEVVMFPITIGLLVASLLSKQYGSFK
                NLRIQFRLQFAHILIPYLIITGITGLFSGFWVLYWHDNNESIHMFVASLLTTLVMHYA
                PSAIKIEYNNNNFTLTPNLQSSPLSNSLQPSPLLSSLQQNDNNNINNNNINNNINNNN
                NNNTNNSGNISDNNNNNINNNNNNNLNENENNQVEIGLNSSTSSTTSSSLSSSTSTII
                NSISINNLSLLENSNLRNSGNGIDTDLTTPLAFEYKIPNQWIIIYSFYIFFMAVNVFL
                LTIPLEDLLLVHFYKGSSIIAFIYLMIEFYIQIKRLRAHYAGGEKSNIKPMKEYFYSV
                YLKFFIYTLVMVAIHYYDKAWRSQKFIHTELAVLVYLVLGVHSIWLLFSPRYYQSDLF
                LPISMLLYFLSNDKERLYLVLFVIPQFYCISNVFYYKLFNVFQPILKQYRKASSTQYK
                DEYKINRLCKSIFNYSLKDINFFSGITDHFLPFIVLLMTSIALSCFKLMDMNFQIGDV
                AIYVPGVYDPPTRPIFSGFVMGFHKLGYFFLLGAFLIKLSNPCPPALVPKYWFSQNTH
                NSNTNNGGENIGDDQERRRKSGGFIFGYKSKSNADFNRIQESLNSQVWGFLMLLLLCS
                LFFFHLGYYNFMITKSFVFTVTISVVVLFYGVSLVCSSIGTLFQNLFFRFILKKRNYI
                HHNLPNIISLDNFTNSNNNNIQYKKINYN"
     CDS        join(20262. .20693,20770. .21645)
                /note="GeneID exon scores (in order of location ranges):
                52.38, 91.33 - GSCJ_ID dd_00782"
                /codon_start=1

Alignment Scores:

Pred. No.:              8.08e+03        Length:         39575
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-858-332G-28 (1-6) x AC116925 (1-39575)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      35567 CATAATCATAATCATAAT 35550

RESULT 428
DMC17A9
LOCUS       DMC17A9                39582 bp    DNA     linear   INV 12-FEB-1998
DEFINITION  Drosophila melanogaster cosmid 17A9.
ACCESSION   AL009146
VERSION     AL009146.1  GI:2827480
KEYWORDS    .
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 39582)
  AUTHORS   Ferraz,C., Vidal,S., Brun,C., Bucheton,A. and Demaille,J.G.
  TITLE     Sequencing the distal X chromosome of Drosophila melanogaster
  JOURNAL   Unpublished
  REMARK    Centre de Recherches de Biochimie Macromoleculaire du CNRS, 1919,
            route de Mende, BP 5051, 34033 MONTPELLIER Cedex, France.
REFERENCE   2  (bases 1 to 39582)
  AUTHORS   Benos,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-FEB-1998) European Drosophila Genome Sequencing
            Consortium
COMMENT     On Feb 1, 1998 this sequence version replaced gi:2653366.
            Sequence submitted by Takis Benos, EMBL Outstation - The EBI,
            Hinxton, Cambridge, CB10 1SD, U.K.
            E-mail: benos@ebi.ac.uk on behalf of the European Drosophila Genome
            Sequencing Consortium. For further information see the European
            Drosophila Genome Sequencing Consortium's web site:
            http://edgp.ebi.ac.uk/.
            Coding sequences are predicted from computer analysis, using both
            gene and CDS prediction programs and matches to other sequences.
            These predictions and matches have been evaluated by the annotators
            and may have been refined by hand. The annotators have also used
            their judgement on what matches to represent in this record. A far
            more complete annotation record is available from FlyBase
            (http://flybase.bio.indiana.edu/) through the FlyBase Annotation
            Object linked by the db_xref qualifier in the Feature Table.
            IMPORTANT:  This sequence is NOT necessarily the entire insert of
            cosmid 17A9.  It may be shorter, since we are minimising the
            overlap between cosmids to 100 bases, by trimming them. The true
            left end of clone 17A9 is at position 19649 of clone 25D2
            Cosmid=17A9; Contig ID=17; Length=39582; Status=Finished.
FEATURES              Location/Qualifiers
     source           1. .39582
                      /organism="Drosophila melanogaster"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:7227"
                      /clone="17A9"
     variation        880
                      /note="variation with respect to cosmid 25D2"
                      /replace="g"
     variation        2017
                      /note="variation with respect to cosmid 25D2"
                      /replace=""
     variation        2881
                      /note="variation with respect to cosmid 25D2"
                      /replace="c"
     variation        2908
                      /note="variation with respect to cosmid 25D2"
```

```
                     /replace=""
variation            3532. .3533
                     /note="variation with respect to cosmid 25D2"
                     /replace="ac"
variation            3849. .3850
                     /note="variation with respect to cosmid 25D2"
                     /replace=""
variation            4161
                     /note="variation with respect to cosmid 25D2"
                     /replace="g"
variation            4696
                     /note="variation with respect to cosmid 25D2"
                     /replace="g"
variation            5303
                     /note="variation with respect to cosmid 25D2"
                     /replace="a"
variation            6473
                     /note="variation with respect to cosmid 25D2"
                     /replace="a"
variation            6492
                     /note="variation with respect to cosmid 25D2"
                     /replace="a"
variation            6587
                     /note="variation with respect to cosmid 25D2"
                     /replace="a"
variation            6763
                     /note="variation with respect to cosmid 25D2"
                     /replace="c"
variation            6894
                     /note="variation with respect to cosmid 25D2"
                     /replace="a"
variation            7104
                     /note="variation with respect to cosmid 25D2"
                     /replace="g"
variation            7488. .7491
                     /note="variation with respect to cosmid 25D2"
                     /replace=""
variation            7711
                     /note="variation with respect to cosmid 25D2"
                     /replace="aag"
variation            7872
                     /note="variation with respect to cosmid 25D2"
                     /replace="c"
variation            7917
                     /note="variation with respect to cosmid 25D2"
                     /replace=""
variation            8552
                     /note="variation with respect to cosmid 25D2"
                     /replace="g"
gene                 8868. .36523
                     /gene="EG:17A9.1"
                     /note="synonym: FBgn0000210;br"
CDS                  join(8868. .9053,12749. .12937,13013. .13466,13556. .13829,
                     13917. .14056,14201. .14251,35709. .36523)
                     /gene="EG:17A9.1"
                     /note="alternatively spliced form"
                     /codon_start=1
                     /protein_id="CAA15628.1"
                     /db_xref="GI:2827483"
                     /db_xref="GOA:Q01295"
                     /db_xref="UniProt/Swiss-Prot:Q01295"
                     /translation="MDDTQHFCLRWNNYQSSITSAFENLRDDEAFVDVTLACEGRSIK
                     AHRVVLSACSPYFRELLKSTPCKHPVILLQDVNFMDLHALVEFIYHGEVNVHQKSLQS
                     FLKTAEVLRVSGLTQQQAEDTHSHLAQIQNLANSGGRTPLNTHTQSLPHPHHGSLHDD
                     GGSSTLFSRQGAGSPPPTAVPSLPSHINNQLLKRMAMMHRSSAAAAAEETSHAFKRLR
                     GSDNSLPLSGAVGSGSNNNSPDLPPLHARSASPQQTPADFSTIKHHNNNNTPPLKEEK
                     RNGPTGNGNSGNGNGNGNGASNGNGISISDKLGSLTPSPLARAGADDVKSEPMDMVCS
                     NNNANANDEHSNDSTGEHDANRSSSGDGGKGSLSSGNDEEIGDGLASHHAAPQFIMSP
                     AENKMFHAAAFNFPNIDPSALLGLNTQLQQSGDLAVSPQGPHSITRSAATSPTSSTSS
                     PPSPPTALISPTSSLKGSLAAAVYSLHSHAHGHVLGHATSPPRPGSVGSSVGSNLCTS
                     TSMGCGVNSGNNSGNNNGNNANNNSNNGNATNNNNNSSSSSTSPGSQATAAGGTVTQA
                     GTPPLPLRMPPPTSGGINEPQECPYCRRTFSCYYSLKRHFQDKHEQSDTLYVCEFCHR
                     RYRTKNSLTTHKSLQHRGSSGMLKRLLKTTAIKHGLVGHGHGHGHVHHPHAHHHALSH
                     PRTSLYDFTSELGQPPPGIQ"
CDS                  join(8868. .9053,12749. .12937,13013. .13466,13556. .13829,
                     13917. .14056,14201. .14251,31318. .31568)
                     /gene="EG:17A9.1"
                     /codon_start=1
                     /protein_id="CAA15626.1"
                     /db_xref="GI:2827481"
                     /db_xref="GOA:Q01295"
                     /db_xref="UniProt/Swiss-Prot:Q01295"
                     /translation="MDDTQHFCLRWNNYQSSITSAFENLRDDEAFVDVTLACEGRSIK
                     AHRVVLSACSPYFRELLKSTPCKHPVILLQDVNFMDLHALVEFIYHGEVNVHQKSLQS
                     FLKTAEVLRVSGLTQQQAEDTHSHLAQIQNLANSGGRTPLNTHTQSLPHPHHGSLHDD
                     GGSSTLFSRQGAGSPPPTAVPSLPSHINNQLLKRMAMMHRSSAAAAAEETSHAFKRLR
                     GSDNSLPLSGAVGSGSNNNSPDLPPLHARSASPQQTPADFSTIKHHNNNNTPPLKEEK
                     RNGPTGNGNSGNGNGNGNGASNGNGISISDKLGSLTPSPLARAGADDVKSEPMDMVCS
                     NNNANANDEHSNDSTGEHDANRSSSGDGGKGSLSSGNDEEIGDGLASHHAAPQFIMSP
                     AENKMFHAAAFNFPNIDPSALLGLNTQLQQSGDLAVSPQGGSTGSLLSGVIVPGGSGG
                     TPSNSSSNNNNNNSNNQQQKVEQQSSPHQLLQQQHHSTPHTNSPQLKQEQPKSGGGSC
                     KSSDLHIAAGSERSLSRSSQGMPDAGGHSATPSPTAAYHKRERERERERERERERERE
                     RSLDHERDLERPGGTGSPPPPPPSHHSHFGQHPLSLLPSHHQLHATHHELSAAAAHHA
                     HAHAHAAHAHALARAGSPMEHHHLLHHRRASLSPSGAVSSASGAGGRGGGAGGPGGPG
                     GSLLSSVRAQDVAQANRLLLPLPLNACHRCDVCGKLLSTKLTLKRHKEQQHLQPLNNA
                     VCNLCHKVFRTLNSLNNHKSIYHRRQKNHHSYFHHGAGVSQAGSPGSRLHQSLSSLSA
                     AAAAANNSVNVGGGSVGGAGGNAVAAAAAAAAAAAELLLSPIVGAAAVAGGTASSTLQ
                     LAAAHQQQQQQSSPGIVKPCMDFL"
CDS                  join(8868. .9053,12749. .12937,13013. .13466,13556. .13829,
                     13917. .14056,19797. .20686)
                     /gene="EG:17A9.1"
                     /note="alternatively spliced form"
                     /codon_start=1
                     /protein_id="CAA15629.1"
                     /db_xref="GI:2827484"
                     /db_xref="GOA:Q01295"
                     /db_xref="UniProt/Swiss-Prot:Q01295"
                     /translation="MDDTQHFCLRWNNYQSSITSAFENLRDDEAFVDVTLACEGRSIK
                     AHRVVLSACSPYFRELLKSTPCKHPVILLQDVNFMDLHALVEFIYHGEVNVHQKSLQS
                     FLKTAEVLRVSGLTQQQAEDTHSHLAQIQNLANSGGRTPLNTHTQSLPHPHHGSLHDD
                     GGSSTLFSRQGAGSPPPTAVPSLPSHINNQLLKRMAMMHRSSAAAAAEETSHAFKRLR
                     GSDNSLPLSGAVGSGSNNNSPDLPPLHARSASPQQTPADFSTIKHHNNNNTPPLKEEK
                     RNGPTGNGNSGNGNGNGASNGNGISISDKLGSLTPSPLARAGADDVKSEPMDMVCS
                     NNNANANDEHSNDSTGEHDANRSSSGDGGKGSLSSGNDEEIGDGLASHHAAPQFIMSP
                     AENKMFHAAAFNFPNIDPSALLGKLTTNATTTTTTINNSITNNNNNNNNNNYDYSLPT
                     KNSNSQKTPSPTTTTLTTPTTTTPTRPTAITSASGICGLNLSTFAANGSSSGGSNGGL
                     SMTALLPQQQQQQQQQHQMSQQQQQQQQQQQQQGNSSSGQQQQPNGILACSTPKANTP
                     TTTQQQMYAAVMAAAASASASTSGSANSSLNNSNSTLNTSGGLNNSASGGDDFRCNPC
                     NKNLSSLTRLKRHIQNVHMRPTKEPVCNICKRVYSSLNSLRNHKSIYHRNLKQPKQEP
                     GVGATQAAANSFYHQQHQQQQLNHHSSS"
variation            9487
                     /gene="EG:17A9.1"
                     /note="variation with respect to cosmid 25D2"

Alignment Scores:
Pred. No.:           8.08e+03        Length:         39582
Score:               42.00           Matches:        6
```

```
Percent Similarity:     100.00%         Conservative:  0
Best Local Similarity:  100.00%         Mismatches:    0
Query Match:            100.00%         Indels:        0
DB:                     2               Gaps:          0

US-09-858-332G-28 (1-6) x DMC17A9 (1-39582)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      37182 CACAACCATAACCACAAC 37199

RESULT 429
AX646691/c
LOCUS       AX646691              39703 bp    DNA     linear   PAT 04-MAR-2003
DEFINITION  Sequence 883 from Patent EP1270724.
ACCESSION   AX646691
VERSION     AX646691.1  GI:28799074
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Suwa,M., Asai,K., Akiyama,Y. and Aburatani,H.
  TITLE     Guanosine triphosphate-binding protein coupled receptors
  JOURNAL   Patent: EP 1270724-A 883 02-JAN-2003;
            National Institute of Advanced Industrial Science and Technology
            (JP) ; Center for Advanced Science and Technology Incubation, Ltd.
            (JP)
FEATURES             Location/Qualifiers
     source          1. .39703
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
     CDS             join(201. .524,3968. .4102,4517. .4687,5787. .5888,6700. .6891,
                     14506. .14748,15377. .15440,16990. .17067,18612. .18762,
                     19258. .19390,20773. .20889,20987. .21167,22295. .22422,
                     23003. .23131,23354. .23497,24275. .24444,39365. .39503)
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD69472.1"
                     /db_xref="GI:28799075"
                     /translation="MGPGAPFARVGWPLPLLVVMAAGVAPVWASHSPHLPRPHSRVPP
                     HPSSERRAVYIGALFPMSGGWPGGQACQPAVEMALEDVNSRRDILPDYELKLIHHDSK
                     VALDMGCDPGQATKYLYELLYNDPIKIILMPGCSSVSTLVAEAARMWNLIVLSYGSSS
                     PALSNRQRFPTFFRTHPSATLHNPTRVKLFEKWGWKKIATIQQTTEVFTSTLDDLEER
                     VKEAGIEITFRQSFFSDPAVPVKNLKVYKERLFGKKYVWFLIGWYADNWFKIYDPSIN
                     CTVDEMTEAVEGHITTEIVMLNPANTRSISNMTSQEFVEKLTKRLKRHPEETGGFQEA
                     PLAYDAIWALALALNKTSGGGGRSGVRLEDFNYNNQTITDQIYRAMNSSSFEGVSGHV
                     VFDASGSRMAWTLIEQLQGGSYKKIGYYDSTKDDLSWSKTDKWIGGSPPADQTLVIKT
                     FRFLSQKLFISVSVLSSLGIVLAVVCLSFNIYNSHVRYIQNSQPNLNNLTAVGCSLAL
                     AAVFPLGLDGYHIGRNQFPFVCQARLWLLGLGFSLGYGSMFTKIWWVHTVFTKKEEKK
                     EWRKTLEPWKLYATVGLLVGMDVLTLAIWQIVDPLHRTIEVPLERRCYGQENEQGSVA
                     IRALGLCIFYGYKGLLLLLGIFLAYETKSVSTEKINDHRAVGMAIYNVAVLCLITAPV
                     TMILSSQQDAAFAFASLAIVFSSYITLVVLFVPKMRRLITRGEWQSEAQDTMKTGSST
                     NNNEEEKSRLLEKENRELEKIIAEKEERVSELRHQLQSRQQLRSRRHPPTPPEPSGGL
                     PRGPPEPPDRLSCDGSRVHLLYKVNPQHCPGNGSVRDKDEAERHNQDEDTGNSHLDLI
                     VGGVFTGESN"
ORIGIN

Alignment Scores:
Pred. No.:              8.11e+03        Length:        39703
Score:                  42.00           Matches:       6
Percent Similarity:     100.00%         Conservative:  0
Best Local Similarity:  100.00%         Mismatches:    0
Query Match:            100.00%         Indels:        0
DB:                     6               Gaps:          0

US-09-858-332G-28 (1-6) x AX646691 (1-39703)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      36716 CATAATCATAATCATAAT 36699

RESULT 430
AB065476/c
LOCUS       AB065476              39703 bp    DNA     linear   PRI 23-JUL-2002
DEFINITION  Homo sapiens gene for seven transmembrane helix receptor, complete
            cds, isolate:CBRC7TM_39.
ACCESSION   AB065476
VERSION     AB065476.1  GI:21928256
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Suwa,M., Sato,T., Okouchi,I., Arita,M., Futami,K., Matsumoto,S.,
            Tsutsumi,S., Aburatani,H., Asai,K. and Akiyama,Y.
  TITLE     Genome-wide discovery and analysis of human seven transmembrane
            helix receptor genes
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 39703)
  AUTHORS   Suwa,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUL-2001) Makiko Suwa, Computational Biology Research
            Center (CBRC), National Institute of Advanced Industrial Science
            and Technology (AIST); 2-41-6 Aomi Koto-ku, Tokyo 135-0064, Japan
            (E-mail:m-suwa@aist.go.jp, URL:http://www.cbrc.jp/,
            Tel:81-3-3599-8080, Fax:81-3-3599-8081)
COMMENT     This sequence is a seven transmembrane helix receptor candidate
            predicted from the whole human genome sequences using our automated
            system that contains programs of gene
            finding(GeneDecoder), sequence search, motif-domain assignment and
            transmembrane helix prediction.
            And the sequence is submitted by the collaborative project between
            [Computational Biology Research Center (CBRC), National Institute
            of Advanced Industrial Science and Technology (AIST)] and [Genome
            Science Division, Research Center for Advanced Science and
            Technology (RCAST), University of Tokyo].
FEATURES             Location/Qualifiers
     source          1. .39703
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /isolate="CBRC7TM_39"
                     /db_xref="taxon:9606"
                     /chromosome="6"
     CDS             join(201. .524,3968. .4102,4517. .4687,5787. .5888,6700. .6891,
                     14506. .14748,15377. .15440,16990. .17067,18612. .18762,
                     19258. .19390,20773. .20889,20987. .21167,22295. .22422,
                     23003. .23131,23354. .23497,24275. .24444,39365. .39503)
                     /codon_start=1
                     /evidence=not_experimental
                     /product="seven transmembrane helix receptor"
                     /protein_id="BAC05730.1"
                     /db_xref="GI:21928257"
                     /translation="MGPGAPFARVGWPLPLLVVMAAGVAPVWASHSPHLPRPHSRVPP
                     HPSSERRAVYIGALFPMSGGWPGGQACQPAVEMALEDVNSRRDILPDYELKLIHHDSK
                     VALDMGCDPGQATKYLYELLYNDPIKIILMPGCSSVSTLVAEAARMWNLIVLSYGSSS
                     PALSNRQRFPTFFRTHPSATLHNPTRVKLFEKWGWKKIATIQQTTEVFTSTLDDLEER
                     VKEAGIEITFRQSFFSDPAVPVKNLKVYKERLFGKKYVWFLIGWYADNWFKIYDPSIN
                     CTVDEMTEAVEGHITTEIVMLNPANTRSISNMTSQEFVEKLTKRLKRHPEETGGFQEA
                     PLAYDAIWALALALNKTSGGGGRSGVRLEDFNYNNQTITDQIYRAMNSSSFEGVSGHV
                     VFDASGSRMAWTLIEQLQGGSYKKIGYYDSTKDDLSWSKTDKWIGGSPPADQTLVIKT
                     FRFLSQKLFISVSVLSSLGIVLAVVCLSFNIYNSHVRYIQNSQPNLNNLTAVGCSLAL
                     AAVFPLGLDGYHIGRNQFPFVCQARLWLLGLGFSLGYGSMFTKIWWVHTVFTKKEEKK
                     EWRKTLEPWKLYATVGLLVGMDVLTLAIWQIVDPLHRTIEVPLERRCYGQENEQGSVA
                     IRALGLCIFYGYKGLLLLLGIFLAYETKSVSTEKINDHRAVGMAIYNVAVLCLITAPV
                     TMILSSQQDAAFAFASLAIVFSSYITLVVLFVPKMRRLITRGEWQSEAQDTMKTGSST
                     NNNEEEKSRLLEKENRELEKIIAEKEERVSELRHQLQSRQQLRSRRHPPTPPEPSGGL
                     PRGPPEPPDRLSCDGSRVHLLYKVNPQHCPGNGSVRDKDEAERHNQDEDTGNSHLDLI
                     VGGVFTGESN"
ORIGIN
```

```
Alignment Scores:
Pred. No.:              8.11e+03        Length:         39703
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     8               Gaps:           0

US-09-858-332G-28 (1-6) x AB065476 (1-39703)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      36716 CATAATCATAATCATAAT 36699

RESULT 431
AL845551/c
LOCUS       AL845551               40374 bp    DNA     linear   VRT 16-APR-2005
DEFINITION  Zebrafish DNA sequence from clone XX-119C3 Contains 3 novel genes
            for alpha-type globins and one novel gene for a beta-type globin,
            complete sequence.
ACCESSION   AL845551
VERSION     AL845551.2  GI:23380956
KEYWORDS    HTG; globin; hemoglobin.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 40374)
  AUTHORS   Matthews,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-DEC-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests:
            http://www.sanger.ac.uk/Projects/D_rerio/faqs.shtml#dataeight
COMMENT     On Sep 30, 2002 this sequence version replaced gi:22416327.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep XX-119C3 is from
            a Incyte Zebrafish AB BAC library VECTOR: pBeloBAC11
            Clone-derived Zebrafish pUC subclones occasionally display
            inconsistency over the length of mononucleotide A/T runs and
            conserved TA repeats. Where this is found the longest good quality
            representation will be submitted.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
            Any regions longer than 1kb tagged as misc-feature 'unsure' are
            part of a tandem repeat of more than 10kb in length where it has
            not been possible to anchor the base differences between repeat
            copies.  The region has been built up based on the repeat element
            to match the total size of repeat indicated by restriction digest,
            but repeat copies may not be in the correct order and the usual
            finishing criteria may not apply.
FEATURES             Location/Qualifiers
     source          1. .40374
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="XX-119C3"
                     /clone_lib="Incyte-BAC"
     misc_feature    1. .48
                     /note="Sequence from uni-directional dGTP big dye
                     terminator reads only."
     misc_feature    1
                     /note="Clone_left_end: XX-119C3"
     gene            complement(3757. .4755)
                     /locus_tag="bY119C3.1-001"
     mRNA            complement(join(3757. .4252,4335. .4542,4624. .4755))
                     /locus_tag="bY119C3.1-001"
                     /product="novel alpha-type globin"
                     /note="match: cDNAs: AB015448.1 AB015449.1 AB026052.1
                     AB080117.1
                     match: ESTs: AL720011.1 AL727420.1 AL721548.1 BQ783706.1
                     AL727315.1 AL727714.1 BE693186.1 AL720600.1 AI617394.1
                     AL719910.1 AL727764.1"
     CDS             complement(join(4124. .4252,4335. .4542,4624. .4718))
                     /locus_tag="bY119C3.1-001"
                     /standard_name="OTTDARP00000001843"
                     /note="match: proteins: Q8UW92 BAC20290 BAC20292 O13169
                     BAC20294 BAC20295"
                     /codon_start=1
                     /product="novel alpha-type globin"
                     /protein_id="CAD60790.1"
                     /db_xref="GI:27802786"
                     /db_xref="GOA:Q7ZT21"
                     /db_xref="InterPro:IPR000971"
                     /db_xref="InterPro:IPR002338"
                     /db_xref="InterPro:IPR002339"
                     /db_xref="UniProt/TrEMBL:Q7ZT21"
                     /db_xref="ZFIN:ZDB-GENE-980526-80"
                     /translation="MSLSAKDKAAVKTLWAKIAGKADDIGHDALSRMLIVYPQTKTYF
                     SHWKDLSPGSAPVRKHGKTVMGGVAEAVSKIDDLNAGLLNLSELHAFQLRVDPANFKI
                     LSHNILVVLATLFPADFTPEAHVAMDKFLSALALAMSEKYR"
     gene            join(7826. .7974,8053. .8816)
                     /locus_tag="bY119C3.2-002"
     gene            join(7826. .7974,8053. .8275,8370. .8816)
                     /locus_tag="bY119C3.2-001"
     mRNA            join(7826. .7974,8053. .8816)
                     /locus_tag="bY119C3.2-002"
                     /product="novel beta-type globin"
                     /note="match: ESTs: AL727267.1"
     mRNA            join(7826. .7974,8053. .8275,8370. .8816)
                     /locus_tag="bY119C3.2-001"
                     /product="novel beta-type globin"
                     /note="match: cDNAs: AF082662.1
                     match: ESTs: BM341906.1 AL909871.1 AI943217.1 AI965114.1
                     BM534785.1"
     CDS             join(7883. .7974,8053. .8275,8370. .8498)
                     /locus_tag="bY119C3.2-001"
                     /standard_name="OTTDARP00000001844"
                     /note="match: proteins: Q98852 O93548 BAC20296"
                     /codon_start=1
                     /product="novel beta-type globin"
                     /protein_id="CAD60792.1"
                     /db_xref="GI:27802788"
                     /db_xref="GOA:O93548"
                     /db_xref="InterPro:IPR000971"
                     /db_xref="InterPro:IPR002337"
                     /db_xref="UniProt/TrEMBL:O93548"
                     /db_xref="ZFIN:ZDB-GENE-030616-7"
                     /translation="MVVWTDFEKATIQDIFAKADYDVIGPQALARCLIVYPWTQRYFA
                     KFGNLYNAAAILGNPMVAAHGKTVLKGLELAVKNMDNIKATYADLSVLHSEKLHVDPD
                     NFRLLADCLTIVVAAQMGAGFTPEVQAAFQKFIAVAVSALGRQYH"
     CDS             join(7883. .7974,8053. .8284)
                     /locus_tag="bY119C3.2-002"
                     /standard_name="OTTDARP00000001845"
                     /codon_start=1
                     /product="novel beta-type globin"
                     /protein_id="CAD60791.1"
                     /db_xref="GI:27802787"
                     /db_xref="GOA:Q7ZYZ5"
                     /db_xref="InterPro:IPR000971"
                     /db_xref="InterPro:IPR002337"
```

```
                     /db_xref="UniProt/TrEMBL:Q7ZYZ5"
                     /db_xref="ZFIN:ZDB-GENE-030616-7"
                     /translation="MVVWTDFEKATIQDIFAKADYDVIGPQALARCLIVYPWTQRYFA
                     KFGNLYNAAAILGNPMVAAHGKTVLKGLELAVKNMDNIKATYADLSVLHSEKLHVDPD
                     NFRVS"
     gene            complement(14921. .15637)
                     /locus_tag="bY119C3.3-001"
     mRNA            complement(join(14921. .15137,15217. .15424,15506. .15637))
                     /locus_tag="bY119C3.3-001"
                     /product="novel alpha-type globin"
                     /note="match: cDNAs: AB015448.1 AB026052.1 AB080117.1
                     match: ESTs: AL721548.1 BQ783706.1 BQ618695.1 AI545160.1
                     AI617394.1 AL719910.1 AL721636.1 AL911459.1"
     CDS             complement(join(15009. .15137,15217. .15424,15506. .15600))
                     /locus_tag="bY119C3.3-001"
                     /standard_name="OTTDARP00000001846"
                     /note="match: proteins: Q90487 BAC20290 P83270 Q8UW95
                     O13169"
                     /codon_start=1
                     /product="novel alpha-type globin"
                     /protein_id="CAD60793.1"
                     /db_xref="GI:27802789"
                     /db_xref="GOA:Q7ZT21"
                     /db_xref="InterPro:IPR000971"
                     /db_xref="InterPro:IPR002338"
                     /db_xref="InterPro:IPR002339"
                     /db_xref="UniProt/TrEMBL:Q7ZT21"
                     /db_xref="ZFIN:ZDB-GENE-980526-80"
                     /translation="MSLSAKDKAAVKTLWAKIAGKADDIGHDALSRMLIVYPQTKTYF
                     SHWKDLSPGSAPVRKHGKTVMGGVAEAVSKIDDLNAGLLNLSELHAFQLRVDPANFKI
                     LSHNILVVLATLFPADFTPEAHVAMDKFLSALALAMSEKYR"
     gene            complement(23127. .24272)
                     /locus_tag="bY119C3.4-001"
     mRNA            complement(join(23127. .23351,23715. .23922,24142. .24272))
                     /locus_tag="bY119C3.4-001"
                     /product="novel alpha-type globin"
                     /note="match: cDNAs: AB015448.1 AB034640.1 AB015449.1
                     match: ESTs: AI477488.1 AL912768.1 AL912769.1 AI353083.1
                     AI353675.1 AI354176.1"
     CDS             complement(join(23223. .23351,23715. .23922,24142. .24236))
                     /locus_tag="bY119C3.4-001"
                     /standard_name="OTTDARP00000001847"
                     /note="match: proteins: BAC20301 Q9YGF7 BAC20294 O13137"
                     /codon_start=1
                     /product="novel alpha-type globin"
                     /protein_id="CAD60794.1"
                     /db_xref="GI:27802790"
                     /db_xref="GOA:Q7ZYZ4"
                     /db_xref="InterPro:IPR000971"
                     /db_xref="InterPro:IPR002338"
                     /db_xref="UniProt/TrEMBL:Q7ZYZ4"
                     /db_xref="ZFIN:ZDB-GENE-990706-3"
                     /translation="MSLSAKDKANVKAFFDKVAPKAEEIGRETLSRTLFVYPQTKTYF
                     SHWADLSPNSPQVKKHGTTVINGVLTAVGLMDDLKGGLLTLSELHAFMLRVDPANFKI
                     INHNLLVSLAMMFPDDFTPEVHVSVDKFLAQVSLALSEKYR"
     misc_feature    40374
                     /note="Clone_right_end: XX-119C3"
ORIGIN

Alignment Scores:
Pred. No.:              8.24e+03        Length:         40374
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     5               Gaps:           0

US-09-858-332G-28 (1-6) x AL845551 (1-40374)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      10229 CATAATCATAATCATAAT 10212
```

```
RESULT 432
SCLACHXII/c
LOCUS       SCLACHXII              40690 bp    DNA     linear   PLN 18-APR-2005
DEFINITION  S.cerevisiae (EU) DNA from left arm of chromosome XII.
ACCESSION   X94607
VERSION     X94607.1  GI:1181264
KEYWORDS    C-5 sterol desaturase; ERG3 gene; FRS1 gene; glycine
            hydroxymethyltransferase; MEF1 gene; NAB1B gene; nucleic acid
            binding protein; PDC1 gene; PET309 gene; phenylalanyl-tRNA
            synthetase alpha subunit; pyruvate decarboxylase; SHMT2 gene; SPT8
            gene; transcription factor; YNT20 gene.
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1
  AUTHORS   Urrestarazu,L.A.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 40690)
  AUTHORS   Urrestarazu,L.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-DEC-1995) L.A. Urrestarazu, Universite Libre de
            Bruxelles, Molecular Biology, Cell Physiology & Yeast Genetics
            Lab., ULB-Campus Plaine-CP244, Bd. du Triomphe - B-1050 Brussels,
            Belgium
COMMENT     Related sequences:
            X04675, D43948, M88277, L19183, U15552, M94955, M62623, M64989,
            U03458, L22529, L47977, J03964, D50617, X60795, L06072, X58378.
FEATURES             Location/Qualifiers
     source          1. .40690
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4932"
                     /chromosome="XII"
                     /map="left arm"
                     /clone_lib="cosmid clone 2C5 (J.Hoheisel, Heidelberg)"
     gene            complement(<1. .417)
                     /gene="PDC1; L2104"
     CDS             complement(<1. .417)
                     /gene="PDC1; L2104"
                     /EC_number="4.1.1.1"
                     /codon_start=1
                     /product="pyruvate decarboxylase"
                     /protein_id="CAA64291.1"
                     /db_xref="GI:1181265"
                     /db_xref="GOA:P06169"
                     /db_xref="UniProt/Swiss-Prot:P06169"
                     /translation="MSEITLGKYLFERLKQVNVNTVFGLPGDFNLSLLDKIYEVEGMR
                     WAGNANELNAAYAADGYARIKGMSCIITTFGVGELSALNGIAGSYAEHVGVLHVVGVP
                     SISAQAKQLLLHHTLGNGDFTVFHRMSANISETTAMI"
     gene            complement(1373. .4039)
                     /gene="L2108"
     CDS             complement(1373. .4039)
                     /gene="L2108"
                     /codon_start=1
                     /protein_id="CAA64292.1"
                     /db_xref="GI:1181266"
                     /db_xref="GOA:P46675"
                     /db_xref="InterPro:IPR000357"
                     /db_xref="SGD:S000004035"
                     /db_xref="UniProt/Swiss-Prot:P46675"
                     /translation="MSGEEEVDYTTLPLEERLTYKLWKARLEAYKELNQLFRNSVGDI
                     SRDDNIQIYWRDPTLFAQYITDSNVVAQEQAIVALNSLIDAFASSSLKNAHNITLIST
                     WTPLLVEKGLTSSRATTKTQSMSCILSLCGLDTSITQSVELVIPFFEKKLPKLIAAAA
                     NCVYELMAAFGLTNVNVQTFLPELLKHVPQLAGHGDRNVRSQTMNLIVEIYKVTGNNS
                     DLLEEILFKKLKPIQVKDLHKLFAKVGDEPSSSKMLFEWEKRELEKKRSQEEEARKRK
                     SILSNDEGEYQIDKDGDTLMGMETDMPPSKQQSGVQIDTFSMLPEETILDKLPKDFQE
                     RITSSKWKDRVEALEEFWDSVLSQTKKLKSTSQNYSNLLGIYGHIIQKDANIQAVALA
                     AQSVELICDKLKTPGFSKDYVSLVFTPLLDRTKEKKPSVIEAIRKALLTICKYYDPLA
                     SSGRNEDMLKDILEHMKHKTPQIRMECTQLFNASMKEEKDGYSTLQRYLKDEVVPIVI
                     QIVNDTQPAIRTIGFESFAILIKIFGMNTFVKTLEHLDNLKRKKIEETVKTLPNFSIA
                     SGSTHSTIETNKQTGPMENKFLLKKSSVLPSKRVASSPLRNDNKSKVNPIGSVASASK
                     PSMVAANNKSRVLLTSKSLATPKNVVANSTDKNEKLIEEYKYRLQKLQNDEMIWTKER
```

```
                QSLLEKMNNTENYKIEMIKENEMLREQLKEAQSKLNEKNIQLRSKEIDVNKLSDRVLS
                LENELRNMEIELDRNKKRNDTNLQSMGTISSYSIPSSTVSSNYGVKSLSSALPFKEEE
                DVRRKEDVNYERRSSESIGDLPHRVNSLNIRPYRKNGTGVSSVSDDLDIDFNDSFASE
                ESYKRAAAVTSTLKARIEKMKAKSRREGTTRT"
gene            complement(4337. .5149)
                /gene="L2111"
CDS             complement(4337. .5149)
                /gene="L2111"
                /codon_start=1
                /protein_id="CAA64293.1"
                /db_xref="GI:1181267"
                /db_xref="GOA:Q12253"
                /db_xref="InterPro:IPR007568"
                /db_xref="SGD:S000004036"
                /db_xref="UniProt/TrEMBL:Q12253"
                /translation="MATHTSKRRIHRWENNELSEENSTIIYFPARGLMWTHFPFVLGI
                CLEFVGYVLKIVFINSPSISTFIAQSVLLLIAPSLYALSIFMLFSKMARLILMEAYML
                IPAKFSTVSFVVADMIGRVLQAVGGGLLSSWNSRNTGRILIIVGLFIQIFCYTFLTFS
                QLFLHYKMKATPSKIVRDSNEWFQYNFILLAGILLVNGRTIVRVVQFLMGLQSYIGQH
                EWCLYVFDTVLMFLLPLIFLATFRARNLFKLQDKSVNIQLNKLLDKESVSED"
gene            complement(5683. .7743)
                /gene="L2114"
CDS             complement(5683. .7743)
                /gene="L2114"
                /codon_start=1
                /protein_id="CAA64294.1"
                /db_xref="GI:1181268"
                /db_xref="GOA:Q12209"
                /db_xref="InterPro:IPR002916"
                /db_xref="SGD:S000004037"
                /db_xref="UniProt/TrEMBL:Q12209"
                /translation="MNLKSIVSWFKEHLPSFDVDVDKHFRTLRVRKYSQICLLISFII
                ICVIIPLMNYLLLTDKFFKICHHLKHHVFNRRSWVHKTHMYHKQSLQLCLICFVFTSF
                FVIQGANGDLLEITKRMGRISVALMPPLLFLTLRPSPLPHTLYLALLPLHKWISRIVV
                LESILHTWFYLYYMYINDTLYVKMRKLPNIYGVIALGLFLLIAITSVRYARRWSYRVF
                YYVHYVSTWLILVFLHYHARPGISYYTTLNVLILTGQIVYRLHITNVTRVTIVPISSS
                LSLLEFPLTDLPKKPILPGGHLRINIYHRNFLRRFFSHLIPFQHPFTIASIPSDNLVR
                LIIRNGHFPLRTNEKYYITGAFEPELSFISKPTVPFNITTKSSKNPFRNNSSALINSP
                LNFLIKAQRVFMCVGGSGISFGLPLLRILNFNGVNVRLLWVSRDYKDLEVLNHFKNNF
                EGMEIYISGTEGNEQDIEIDYIDYHDCAADINDEVRSISSSGRVSELGDNSMLSDGNP
                QPTEPNENTALLSKKSTLRNHHPPKTSDIPDINADDEIDFTYAFSRSKSRKNTAQGTL
                TTHSSFNGSSVFRQPKIIEPPAQDPCLEAAPKKIRIPAGVKVFFGRPTLGDKDYEWCL
                QTECDAETDSIQCCRWANQGRDHAEYLSQVWVLAAGPRGLIESTKRWATDGGLHFHGE
                SFAL"
gene            8568. .9685
                /gene="NAB1B; L2118"
CDS             join(8568. .8657,9017. .9685)
                /gene="NAB1B; L2118"
                /note="duplicate on chromosome VII"
                /codon_start=1
                /product="nucleic acid binding protein"
                /protein_id="CAA64295.1"
                /db_xref="GI:1181269"
                /db_xref="GOA:P46654"
                /db_xref="SGD:S000004038"
                /db_xref="UniProt/Swiss-Prot:P46654"
                /translation="MSLPATFDLTPEDAQLLLAANTHLGARNVQVHQEPYVFNARPDG
                VHVINVGKTWEKLVLAARIIAAIPNPEDVVAISSRTYGQRAVLKFAAHTGATPIAGRF
                TPGSFTNYITRSFKEPRLVIVTDPRLDAQAIKEASYVNIPVIALTDLDSPSEFVDVAI
                PCNNRGKHSIGLIWYLLAREVLRLRGALVDRTQPWSIMPDLYFYRNPEEVEQVAEEAA
                AAEEGEEEEVKEEVTEGQAEATEWAEENADNVEW"
exon            8568. .8657
                /gene="NAB1B; L2118"
                /number=1
intron          8658. .9016
                /gene="NAB1B; L2118"
                /number=1
exon            9017. .9685
                /gene="NAB1B; L2118"
                /number=2
gene            complement(10222. .11508)
                /gene="L2121"
CDS             complement(10222. .11508)
                /gene="L2121"
                /codon_start=1
                /protein_id="CAA64296.1"
                /db_xref="GI:1181270"
                /db_xref="SGD:S000004039"
                /db_xref="UniProt/TrEMBL:Q12110"
                /translation="MVNNIMHEYVPPSQRLFHSRHRITRNDLKEEALHSGTTDWTTIL
                DTTIDKDTNLISYAVPIIDNFAIPRANSQGSPVAPSPNHRSTMYSSSSSSASSVFSDG
                LFTPNNNRNSSGSSSLVIRPQKNLSVDSLIQENKRKINSEKESLSLIANNNDETLCTH
                TDPSIQNLIKSETKRNILNLKFQNRNLFRRELKLEKFWSNLRSCHTSGDETDLLLVIS
                KHNLYWFGIPNDFRLPIYKRCLYHYSELDEAGFFSQYANNSLYLAIRKCCNNEEQETL
                SRSIFINLTKNVTWLNSRFDDNKDNKNSYTVTEGKFYQDFPNLYYHLKDKLKLNVIMD
                FIKPVIRNFMTNALNKHKLDGIGLELLDILIVTTYYGPNKINAFLMDTFILNLLKQCH
                YKFFVSNISELVIQISKIDCDLVILLEDLRSRIDLD"
gene            complement(11923. .12408)
                /gene="L2125"
CDS             complement(11923. .12408)
                /gene="L2125"
                /codon_start=1
                /protein_id="CAA64297.1"
                /db_xref="GI:1181271"
                /db_xref="GOA:Q12155"
                /db_xref="SGD:S000004040"
                /db_xref="UniProt/TrEMBL:Q12155"
                /translation="MKLGHREQQFYLWYFIVHIPITIFIDSSVVIPAKWQLGIAQKVV
                SDHIAKQHDFLLSEKPEWLYWFVVLELVLQLPLFVYFVNKFWNSSELQVNTNSRLKKW
                LRIYGWNASLTTLICIVVIFKRGYIPYDVLKTSLSMTQKCQLASVYLPTFLIPLRLCF
                V"
gene            complement(12660. .13313)
                /gene="L2128"
CDS             complement(12660. .13313)
                /gene="L2128"
                /codon_start=1
                /protein_id="CAA64298.1"
                /db_xref="GI:1181272"
                /db_xref="GOA:Q12035"
                /db_xref="SGD:S000004041"
                /db_xref="UniProt/TrEMBL:Q12035"
                /translation="MDQSVEDLFGALRDASASLEVKNSAKEQVSLQQEDVLQIGNNDD
                EVEIESKFQEIETNLKKLPKLETGFDALANKKKKKNVLPSVETEDKRKPNKSDKNDND
                WFTLPKPDDNMRREVQRDLLLIKHRAALDPKRHYKKQRWEVPERFAIGTIIEDKSEFY
                SSRMNRKERKSTILETLMGDEASNKYFKRKYNEIQEKSTSGRKAHYKKMKEMRKKRR"
gene            13537. .14289
                /gene="L2131"
CDS             13537. .14289
                /gene="L2131"
                /codon_start=1
                /protein_id="CAA64299.1"
                /db_xref="GI:1181273"
                /db_xref="GOA:Q12345"
                /db_xref="SGD:S000004042"
                /db_xref="UniProt/Swiss-Prot:Q12345"
                /translation="MKFEDLLATNKQVQFAHAATQHYKSVKTPDFLEKDPHHKKFHNA
                DGLNQQGSSTPSTATDANAASTASTHTNTTTFKRHIVAVDDISKMNYEMIKNSPGNVI
                TNANQDEIDISTLKTRLYKDNLYAMNDNFLQAVNDQIVTLNAAEQDQETEDPDLSDDE
                KIDILTKIQENLLEEYQKLSQKERKWFILKELLLDANVELDLFSNRGRKASHPIAFGA

Alignment Scores:
Pred. No.:              8.3e+03         Length:         40690
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x SCLACHXII (1-40690)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      17590 CATAACCATAATCATAAT 17573

RESULT 433
DMC140G11/c
```

```
LOCUS       DMC140G11              40933 bp    DNA     linear   INV 16-APR-2005
DEFINITION  Drosophila melanogaster cosmid clone 140G11.
ACCESSION   AL035395
VERSION     AL035395.1  GI:4454123
KEYWORDS    .
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 40933)
  AUTHORS   Borkova,D., Minana,B. and Kafatos,F.C.
  TITLE     Sequencing the distal X chromosome of Drosophila melanogaster
  JOURNAL   Unpublished
  REMARK    European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69117,
            Heidelberg, Germany.
REFERENCE   2  (bases 1 to 40933)
  AUTHORS   Benos,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-DEC-1999) European Drosophila Genome Sequencing
            Consortium
COMMENT     Clone=140G11; Contig ID=26; Length=40933; Status=Finished Sequence
            submitted by Takis Benos, EMBL Outstation - The EBI, Hinxton,
            Cambridge, CB10 1SD, U.K.
            E-mail: benos@ebi.ac.uk on behalf of the European Drosophila Genome
            Sequencing Consortium. For further information see the European
            Drosophila Genome Sequencing Consortium's web site:
            http://edgp.ebi.ac.uk/.
            Coding sequences are predicted from computer analysis, using both
            gene and CDS prediction programs and matches to other sequences.
            These predictions and matches have been evaluated by the annotators
            and may have been refined by hand (in which case a Genefinder
            prediction will have no score. The annotators have also used their
            judgement on what matches to represent in this record. A far more
            complete annotation record is available from FlyBase
            (http://flybase.bio.indiana.edu/) through the FlyBase Annotation
            Object linked by the db_xref qualifier in the Feature Table.
            IMPORTANT:  This sequence is NOT necessarily the entire insert of
            clone 140G11.  It may be shorter, since we are minimising the
            overlap between clones to 100 bases, by trimming them. Clone 140G11
            overlaps to the left with clone 163A10 This is the rightmost clone
            of its contig.
            The syntax for the representation of annotation used in this record
            is documented at:
            ftp://ftp.ebi.ac.uk/pub/databases/edgp/sequence_annotation.README.
FEATURES             Location/Qualifiers
     source          1. .40933
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="cosmid 140G11"
     gene            join(<1993. .8141,8236. .8545,8618. .9076,9157. .9503)
                     /gene="EG:163A10.2"
                     /note="synonym: FBgn0004647;N"
     mRNA            join(<1993. .8141,8236. .8545,8618. .9076,9157. .9503)
                     /gene="EG:163A10.2"
     gene            complement(11199. .11498)
                     /gene="EG:140G11.3"
     CDS             complement(11199. .11498)
                     /gene="EG:140G11.3"
                     /note="/prediction=(method:''genscan'', version:''1.0'',
                     score:''20.04'')
                     /prediction=(method:''genefinder'', version:''084'')
                     /match=(desc:''GH06477.5prime GH Drosophila melanogaster
                     head pOT2 Drosophila melanogaster cDNA clone GH06477
                     5prime similar to M16152: D.melanogaster growth factor
                     (Notch locus) gene, exon 6, mRNA sequence'',
                     species:''Drosophila melanogaster (fruit fly)'',
                     ranges:(query:11104. .11514, target:EMBL::AI107162:453. .43,
                     score:''1830.00'')), method:''blastn'',
                     version:''1.4.9'')"
                     /codon_start=1
                     /protein_id="CAB37612.1"
                     /db_xref="GI:6691843"
                     /db_xref="FLYBASE:FBgn0028746"
                     /db_xref="UniProt/TrEMBL:O97172"
                     /translation="MVCVPCIIIPLLLYIWHKFVQPILLRYWNPWEKKDDDGNVIKKG
                     PDFPFECKGGVCPFVPGGKKTENVSDDDAEESENPPLNATAMAAETEVDESKKEI"
     gene            complement(join(11928. .12113,12186. .12641))
                     /gene="EG:140G11.2"
                     /note="synonym: FBgn0000644;Fcp3C"
     CDS             complement(join(11928. .12113,12186. .12641))
                     /gene="EG:140G11.2"
                     /codon_start=1
                     /protein_id="CAB37613.1"
                     /db_xref="GI:6691844"
                     /db_xref="UniProt/Swiss-Prot:P11450"
                     /translation="MSSTGVASSSTTAEEDWPTAVEFVIMTTPASELEASTETIGNNG
                     TTETTVGEAPIIGSSEGSTRSMEPTTASPLMSTNPSSSSSLVSTIPLPPTAGLPVQDN
                     QPVPCTCGVFLSSQIPNGLPTKPLIHQELDHMFPCNAIGRKQCQTKCLETIVQHLPNS
                     ANIVCSALGHDCHKERAYLFIKNCHNQWVNTNLQAGREYCCRSGVPYRCPLMG"
     gene            join(14042. .14186,14462. .14724)
                     /gene="EG:140G11.5"
     CDS             join(14042. .14186,14462. .14724)
                     /gene="EG:140G11.5"
                     /note="/prediction=(method:''genefinder'',
                     version:''084'', score:''23.69'')
                     /prediction=(method:''genscan'', version:''1.0'',
                     score:''39.63'')
                     /match=(desc:''SD07039.5prime SD Drosophila melanogaster
                     Schneider L2 cell culture pOT2 Drosophila melanogaster
                     cDNA clone SD07039 5prime, mRNA sequence'',
                     species:''Drosophila melanogaster (fruit fly)'',
                     ranges:(query:14007. .14186,
                     target:EMBL::AI534269:249. .428, score:''891.00''),
                     (query:14459. .14618, target:EMBL::AI534269:426. .585,
                     score:''782.00'')), method:''blastn'', version:''1.4.9'')"
                     /codon_start=1
                     /protein_id="CAA23057.1"
                     /db_xref="GI:4454131"
                     /db_xref="FLYBASE:FBgn0040396"
                     /db_xref="GOA:Q9U1K7"
                     /db_xref="InterPro:IPR006662"
                     /db_xref="InterPro:IPR010357"
                     /db_xref="UniProt/TrEMBL:Q9U1K7"
                     /translation="MPEYVPARGFKEMENLIKLYENQRSPIYIYFYGEKDKDGRSWCP
                     DCVAAEETIMSAFRNHAPADCMILVVDVGSRESWIGKDNMFRKPPYSVEGIPTLIRWK
                     GVERLDGDQLLKSSLLELFFEETDPKKSNFVAQ"
     gene            20108. .20755
                     /gene="EG:140G11.4"
                     /note="synonym: FBgn0000479;dnc"
     mRNA            20108. .20755
                     /gene="EG:140G11.4"
ORIGIN

Alignment Scores:
Pred. No.:                8.35e+03        Length:         40933
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       2               Gaps:           0

US-09-858-332G-28 (1-6) x DMC140G11 (1-40933)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      18970 CACAATCACAATCACAAT 18953

RESULT 434
U41559
LOCUS       U41559                 41511 bp    DNA     linear   INV 25-OCT-2004
DEFINITION  Caenorhabditis elegans cosmid C26B2, complete sequence.
ACCESSION   U41559
VERSION     U41559.2  GI:15145284
KEYWORDS    HTG.
```

```
SOURCE      Caenorhabditis elegans
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 41511)
  AUTHORS   .
  CONSRTM   WormBase Consortium
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
  JOURNAL   Science 282 (5396), 2012-2018 (1998)
   PUBMED   9851916
REFERENCE   2  (bases 1 to 41511)
  AUTHORS   Du,Z. and Gattung,S.
  TITLE     The sequence of C. elegans cosmid C26B2
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 41511)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-NOV-1995) Robert Waterston
REFERENCE   4  (bases 1 to 41511)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUN-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   5  (bases 1 to 41511)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-2001) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   6  (bases 1 to 41511)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAY-2002) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   7  (bases 1 to 41511)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2002) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   8  (bases 1 to 41511)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-2002) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   9  (bases 1 to 41511)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2003) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   10 (bases 1 to 41511)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-2003) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   11 (bases 1 to 41511)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-DEC-2003) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   12 (bases 1 to 41511)
  AUTHORS   .
  CONSRTM   WormBase Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (16-SEP-2004) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
REFERENCE   13 (bases 1 to 41511)
  AUTHORS   .
  CONSRTM   WormBase Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-2004) Department of Genetics, Washington
            University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
            Louis, MO 63110, USA
COMMENT     On Aug 9, 2001 this sequence version replaced gi:1118107.
            Submitted by:
                  Genome Sequencing Center
                  Department of Genetics, Washington University
                  St. Louis , MO  63110, USA, and
                  Sanger Centre, Hinxton Hall
                  Cambridge CB10 IRQ, England
                  email: submissions@watson.wustl.edu and jes@sanger.ac.uk

            NOTICE:  This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate chemistry
            or covered by high quality data (i.e., phred quality >= 30); an
            attempt was made to resolve all sequencing problems, such as
            compressions and repeats; all regions were covered by sequence from
            more than one m13 subclone.

            For a graphical representation of this clone sequence and its
            analysis see:
            http://www.wormbase.org/db/seq/sequence?name=C26B2;class=Sequence

                        NEIGHBORING CLONE INFORMATION

            The 5' clone is Y42H9B, 200 bp overlap; the 3' clone is C18F3, 200
            bp overlap. Actual start of this clone is at base position 1 of
            C26B2; actual end is at 41511 of C26B2.

            NOTES:

            Coding seqences below are the result of integration and manual
            review of the following data : computer analysis using the program
            Genefinder (P. Green and L. Hillier, personal communication), the
            large scale EST projects of Yuji Kohara
            (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C.
            elegans ORFeome cloning project (http://worfdb.dfci.harvard.edu/),
            similarity to other proteins from BlastX analyses
            (http://blast.wustl.edu/), sequence conservation with C. briggsae
            using Jim Kent's WABA alignment program (Genome Research
            10:1115-1125, 2000), individual C. elegans GenBank submissions,
            and personal communications with C. elegans researchers. tRNAs
            are predicted using the program tRNAscan-SE (Lowe, T.M. and
            Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).
FEATURES                Location/Qualifiers
     source             1. .41511
                        /organism="Caenorhabditis elegans"
                        /mol_type="genomic DNA"
                        /strain="Bristol N2"
                        /db_xref="taxon:6239"
                        /chromosome="IV"
                        /clone="C26B2"
     gene               1060. .2569
                        /gene="C26B2.8"
                        /locus_tag="C26B2.8"
     CDS                join(1060. .1285,1334. .1455,1501. .1594,1707. .1865,
                        1910. .1971,2015. .2150,2201. .2329,2376. .2569)
                        /gene="C26B2.8"
                        /locus_tag="C26B2.8"
                        /standard_name="C26B2.8"
                        /note="contains similarity to Interpro domains IPR008914
                        (PEBP), IPR006150 (Worm-specific repeat type 1); coded for
                        by the following C. elegans cDNAs: OSTF137G7_1,
                        OSTR137G7_1"
```

```
                     /codon_start=1
                     /product="Hypothetical protein C26B2.8"
                     /protein_id="AAC24260.1"
                     /db_xref="GI:1118115"
                     /db_xref="WormBase:C26B2.8"
                     /translation="MCWASCLRADLDWHVNINKRFLLEVRPWAPRYINHREMLHFIAV
                     FLICFLPAIDNMLNTVFIMLLLPAAGLCARYNSHLVKGWSNSDCVRPWQQPLMSKPTL
                     RIPSRAIECHPEEHDCPTGRNAVCQYSLRSQKYVCCEDKKDADIPTCPKFYETLLLPC
                     GNSVDSQCPRGYRCLGSLGDDSVKLCCKPNRTLEYREPEHTFRENRIVPRLLPVAPAY
                     ELIATFSDEQISMGQLFDADILDKLADPPVMSAGVELQDEKLYTIILADATSKAVTWL
                     VANIAAFDGQLEIHRRTKSAVSYQPPDSTDRPVGMHTMILALYEQNDTWTQKDLARIA
                     MDDFNFTEWLEEYAHVLPAQALAATFYGYSTKKDDRKRI"
     gene            3212. .3723
                     /gene="C26B2.7"
                     /locus_tag="C26B2.7"
     CDS             join(3212. .3312,3363. .3465,3520. .3723)
                     /gene="C26B2.7"
                     /locus_tag="C26B2.7"
                     /standard_name="C26B2.7"
                     /note="contains similarity to Arabidopsis thaliana
                     Hypothetical protein At2g01930 (Expressed protein).;
                     TR:Q9SKD0; coded for by the following C. elegans cDNAs:
                     OSTF144F2_1, OSTR144F2_1, yk651e1.3, yk651e1.5,
                     yk1569e08.3, yk1569e08.5"
                     /codon_start=1
                     /product="Hypothetical protein C26B2.7"
                     /protein_id="AAC24261.1"
                     /db_xref="GI:1118114"
                     /db_xref="WormBase:C26B2.7"
                     /translation="MLRIIQRIPTERYLCTLPVKNPKLHRTKDQRDITNMDLLEEQIR
                     IKSAEKNREMTIQGQKVKWAYRSELVGRHDLDKNHRNTWMAYLIIILVGFGAFVTVKS
                     QVVMGRRQEMDAREALRKELQLSGDDRKKIGVV"
     misc_feature    3934. .3935
                     /note="SL2 trans-splice acceptor; see yk1341f02.5"
     gene            3936. .5065
                     /gene="C26B2.6"
                     /locus_tag="C26B2.6"
     CDS             join(3936. .4071,4147. .4557,4608. .5065)
                     /gene="C26B2.6"
                     /locus_tag="C26B2.6"
                     /standard_name="C26B2.6"
                     /note="contains similarity to Pfam domain PF05625 (PAXNEB
                     protein); coded for by the following C. elegans cDNAs:
                     OSTF025A8_1, OSTR025A8_1, yk1384a01.3"
                     /codon_start=1
                     /product="Hypothetical protein C26B2.6"
                     /protein_id="AAC24262.1"
                     /db_xref="GI:1118113"
                     /db_xref="WormBase:C26B2.6"

Alignment Scores:
Pred. No.:                8.46e+03        Length:         41511
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       2               Gaps:           0

US-09-858-332G-28 (1-6) x U41559 (1-41511)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       9511 CATAATCATAATCATAAT 9528

RESULT 435
NCB12N19
LOCUS       NCB12N19               42273 bp    DNA     linear   PLN 18-JAN-2002
DEFINITION  Neurospora crassa DNA linkage group II BAC clone B12N19.
ACCESSION   AL669987
VERSION     AL669987.1  GI:18376315
KEYWORDS    .
SOURCE      Neurospora crassa
  ORGANISM  Neurospora crassa
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
REFERENCE   1
  AUTHORS   Schulte,U., Aign,V., Hoheisel,J., Brandt,P., Fartmann,B.,
            Holland,R., Nyakatura,G., Mewes,H.W. and Mannhaupt,G.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 42273)
  AUTHORS   German Neurospora genome,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JAN-2002) MIPS, Institut fuer Bioinformatik,
            GSF-Forschungszentrum fuer Umwelt und Gesundheit, GmbH,
            Ingolstaedter Landstrasse 1, D-85764 Neuherberg, FRG, E-mail:
            G.Mannhaupt@gsf.de Project Coordinator: Ulrich Schulte, Institute
            of Biochemistry, Heinrich-Heine-University, D-40225 Duesseldorf,
            E-mail: ulrich-schulte@uni-duesseldorf.de
COMMENT     BAC clone 12N19 (strain OR74A) is available at the Fungal Genetic
            Stock Center, http://www.fgsc.net
            Sequencing was performed by MWG Biotech AG, Ebersberg, Germany,
            http://www.mwgdna.com
            Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of linkage groups II
            and V can be viewed at: http://mips.gsf.de/proj/neurospora.
FEATURES             Location/Qualifiers
     source          1. .42273
                     /organism="Neurospora crassa"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5141"
                     /chromosome="6"
     misc_feature    1. .7212
                     /note="overlap to BAC B11E6, EMBL:AL442043, please refer
                     to this entry for analysis and annotation"
     gene            9830. .11707
                     /gene="B12N19.020"
     CDS             join(9830. .10364,10432. .11357,11450. .11707)
                     /gene="B12N19.020"
                     /note="differs in three residues when compared to
                     PIR:T47219
                     Contains Amino acid permeases signature
                     [GGPASLLIDFAIIGVMMFNVVYALGELAVMFP]
                     contains EST gb:BF739557, BF739423"
                     /codon_start=1
                     /product="amino acid permease NAAP1"
                     /protein_id="CAD21063.1"
                     /db_xref="GI:18376316"
                     /db_xref="GOA:Q8X0I5"
                     /db_xref="UniProt/TrEMBL:Q8X0I5"
                     /translation="MSDIEKGSAGSHDKETRVDAATLPAYEAGSVAEAQQEDFMTRNG
                     LNLASFKPRRFHTGDGTEIERTMKSRHLHMIAIGGSIGAGFFVGSGKALHNGGPASLL
                     IDFAIIGVMMFNVVYALGELAVMFPVTGGFYTYSTRFIDPSWGFAMGWNYVFQWAIVL
                     PLELTVCGFTVGYWNQDITVWITIFLVAIIIINVFGTLGYAEEEFWASAFKLGATIVF
                     LIICLVMVLGGGPKGSRYDHYTGAALWYSPGAFANGFRGFCSCFVTAAFAFSGTELVG
                     LAAAAEAKNPVKSLPGAIKQVFWRITLFYILGLFFVGLLIPYDDENLLGANPFINVNAS
                     PFVLVGKYANLTGFDSFMNVVILVSVLSLGVSCVYGGSRTLTALAQQGYAPKIFTYVD
                     RSNRPLFSVLLIIAFGPLAYVNLAAAGPTVFDWLLSLSGLAALFTWGSICLSHIRFRK
                     AWALKGHSVEEIPFRAAGGVWGSWLGLVLVVLVLIAQFYVAICPPGGGFNTVEGFFVQ
                     YLALPVVIVFWIGGFIWKRTSWLRVDQIDVDTGRREHDWESINAYQAELAAMPAWKRW
                     FHKLFV"
     exon            9830. .10364
                     /gene="B12N19.020"
                     /number=1
     intron          10365. .10431
                     /gene="B12N19.020"
                     /number=1
     exon            10432. .11357
                     /gene="B12N19.020"
                     /number=2
     intron          11358. .11449
                     /gene="B12N19.020"
                     /number=2
     repeat_region   11367. .11389
                     /note="23 bp poly(t) tandem repeat"
     exon            11450. .11707
                     /gene="B12N19.020"
```

```
                /number=3
repeat_region   13995. .14030
                /note="36 bp atatatatatatatatatatatat tandem repeat"
gene            14175. .14909
                /gene="B12N19.030"
CDS             14175. .14909
                /gene="B12N19.030"
                /note="contains EST gb:BG279737, BG279736"
                /codon_start=1
                /product="putative protein"
                /protein_id="CAD21064.1"
                /db_xref="GI:18376317"
                /db_xref="UniProt/TrEMBL:Q8X0I4"
                /translation="MGIFDYLRPRRQNGIPRKHTMPPPDILSYQNDALDKRHSKLSSM
                SPNAEQSQANTRPTTSSSWTSAITSTTTSTSISTSTSTSILTPTTTSHTTTTTPLSPP
                HSNYTLSNHDLNLNRPPTPPSRHPPPPSPPPSPPLFPTSEFTLHLPPPPPPPYARLSS
                RRTDLPLPSYAELVQERQIVPPRPRPLPTPREPPPSFEAAEIQEHDELHGMDPSSWPG
                CEYCRVYELWVSYWRGEGVGGVGVSG"
exon            14175. .14909
                /gene="B12N19.030"
                /number=1
gene            15344. .16676
                /gene="B12N19.040"
gene            complement(join(15344. .16586,16639. .16676))
                /gene="B12N19.040"
CDS             complement(join(15344. .16586,16639. .16676))
                /gene="B12N19.040"
                /note="similarity to heterokaryon incompatibility protein
                het-6, Neurospora crassa, TREMBL:AF208542_1"
                /codon_start=1
                /product="related to heterokaryon incompatibility protein
                het-6"
                /protein_id="CAD21065.1"
                /db_xref="GI:18376318"
                /db_xref="UniProt/TrEMBL:Q8X0I3"
                /translation="MEAHEGGERVGACNSGPVQTTGERYPYRQLDHENTIRLLFITPG
                SGEEASSTYSLLHAELGIIPYQALSYEWGLPSDDDPNITIDGHTVRIRKNLSEALKQI
                SSVIRYLDFLLLWIDALCINQNDDAEKSQQVREMGKIFSSAEQVLAWTGPMANDSDYA
                IDMLNAAPQDRIFDTQLPENYRARTAILAWCNRPYWKRIWIIQELFLAKRLVFMCGSR
                SISRRGNVVANCLWVILKTSEKSLPVVKSDNAEFWDTVLYRGKNNSYTLRDWLYVCAT
                SQIMASDTRDYIYAVLSISHETRSSLAQIIPDYTKSAETVFEDLNTLMMDERAREGCE
                SCLGKVKTFDYIRNKTINRWLIKTLGLNVQIGWHDESQYPSCNSCGRPEREYSKGELS
                AVKVRSEGKEFWTSEVHSLNFTAMVVENLWQGGG"
exon            complement(15344. .16586)
                /gene="B12N19.040"
                /number=1
intron          complement(16587. .16638)
                /gene="B12N19.040"
                /number=1
exon            complement(16639. .16676)
                /gene="B12N19.040"
                /number=2
gene            17463. .18879
                /gene="B12N19.050"
gene            complement(join(17463. .18741,18857. .18879))
                /gene="B12N19.050"
CDS             complement(join(17463. .18741,18857. .18879))
                /gene="B12N19.050"
                /note="similarity to heterokaryon incompatibility protein
                het-6, Neurospora crassa, TREMBL:AF208542_1"
                /codon_start=1
                /product="related to heterokaryon incompatibility protein
                het-6"
                /protein_id="CAD21066.1"
                /db_xref="GI:18376319"
                /db_xref="UniProt/TrEMBL:Q8X0I2"
                /translation="MPPLSYYTIGEDQGEDELPPAILPLTEEGGNAIRLLTLREPSSH
                EGETLSLEMHLSTTSEKYIALSYEWGQAEPDDPMILVNNRQVQIRKNLHEALRQISEH
                YSFYRKQRHDFKCPHCHGSVVDACKCGNRPEKDCHEKSLGPWTNLGDSSFWAKHDPPR
                LWIDALCINQNDNKEKSSQVGRMGTIFSSAKVVLSWIGLPRDGSNDALRMMRTLSMSY
                PELPEFVATGGLTDSLILSIASLCERTYWRRVWILQELYLAQKYVVMCGPESIASDSF
                DEALGQLLDLSLRPFLKSYNIMSRSAANAMVVSKRVSGLCSLVRWLIVVVNADFQATV
                PHDYIYAILAISSPGVIDMLKPGIDTTQIIVDYNKTVAEVYRQLLQSFEAGPSLMNHR
                RWLLKLALKMGLSQEEGVALLDSEHSKRCEVAKGSKSDLDY"
exon            complement(17463. .18741)
                /gene="B12N19.050"
                /number=1
intron          complement(18742. .18856)
                /gene="B12N19.050"
                /number=1
exon            complement(18857. .18879)
                /gene="B12N19.050"
                /number=2
gene            20541. .21062
                /gene="B12N19.060"
gene            complement(20541. .21062)
                /gene="B12N19.060"
CDS             complement(20541. .21062)
                /gene="B12N19.060"
                /note="weak similarity to proteinase umuD (EC 3.4.21.-) -
                Salmonella typhimurium, PIR:A36713"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="CAD21067.1"
                /db_xref="GI:18376320"
                /db_xref="UniProt/TrEMBL:Q8X0I1"
                /translation="MPTVMGLTTDSSSKTTTTLGNQDLKKPNGGSRLDTISKRNHHAA
                KEIETLFWRCLCDDPSAAPEYMAEDCILINPLLSSHPSTTPLSKDSEPNILEVFEKAE
                PWAGFRFHGDPLVVEIDLMAVALVYKVSLYTVDKKGKGTREVVASVSSSWRQTAGADW
                VLVAHHCQYVDDE"
exon            complement(20541. .21062)
                /gene="B12N19.060"
                /number=1
repeat_region   21827. .21854
                /note="28 bp poly(a) tandem repeat"
gene            23273. .29225
                /gene="B12N19.070"
gene            complement(join(23273. .24898,24975. .28445,28506. .28670,
                28743. .29225))
                /gene="B12N19.070"
CDS             complement(join(23273. .24898,24975. .28445,28506. .28670,
                28743. .29225))
                /gene="B12N19.070"
                /note="similarity to regulator of PHR1, YER169w,
                Saccharomyces cerevisiae, PIR:S50672"
                /codon_start=1
                /product="related to regulator of deoxyribodipyrimidine
                photo-lyase PHR1"
                /protein_id="CAD21068.1"
                /db_xref="GI:18376321"

Alignment Scores:
Pred. No.:              8.61e+03        Length:         42273
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x NCB12N19 (1-42273)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      23047 CATAACCATAACCATAAC 23064

RESULT 436
AC006277/c
LOCUS       AC006277               42642 bp    DNA     linear   PRI 05-JAN-1999
DEFINITION  Homo sapiens chromosome 19, cosmid F25540, complete sequence.
ACCESSION   AC006277
VERSION     AC006277.1  GI:4096053
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
```

```
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 42642)
  AUTHORS   Lamerdin,J.E., McCready,P.M., Skowronski,E., Viswanathan,V.,
            Burkhart-Schultz,K., Gordon,L., Dias,J., Ramirez,M., Stilwagen,S.,
            Phan,H., Velasco,N., Do,L., Regala,W., Terry,A., Garnes,J.,
            Danganan,L., Erler,A., Christensen,M., Georgescu,A., Avila,J.,
            Liu,S., Attix,C., Andreise,T., Trankheim,M., Amico-Keller,G.,
            Coefield,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G.,
            Kronmiller,B., Arellano,A., Sanders,C., Ow,D., Nolan,M., Trong,S.,
            Kobayashi,A., Olsen,A.S. and Carrano,A.V.
  TITLE     Sequence analysis of a 3.5 Mb contig in human 19p13.3 containing a
            serine protease gene cluster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 42642)
  AUTHORS   Lamerdin,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-1999) Joint Genome Institute, Lawrence Livermore
            National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
COMMENT     Map and sequence oriented from p telomere to centromere. Cosmid
            F25540 overlaps cosmid R33777 to the left from bases 1 to 552 of
            this accession, and overlaps cosmid R26610 to the right from bases
            39,967 to 42,642. Additional map and sequence information are
            available at: http://www-bio.llnl.gov/bbrp/genome/genome.html.
FEATURES             Location/Qualifiers
     source          1. .42642
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /map="19p13.3 between CDC34 and D19S342"
                     /clone="F25540"
                     /cell_line="UV5HL9-5B"
                     /clone_lib="LL19NC02 F chromosome 19-specific cosmid
                     library"
                     /note="Cosmid library constructed at LLNL from flow-sorted
                     chromosomes from human-hamster hybrid UV5HL9-5B, which
                     carries chromosome 19 as its only human chromosome."
     repeat_region   343. .647
                     /rpt_family="AluJo"
     repeat_region   652. .961
                     /rpt_family="AluSp"
     repeat_region   complement(1204. .1505)
                     /rpt_family="AluSx"
     repeat_region   2124. .2424
                     /rpt_family="AluY"
     misc_feature    complement(2725. .2856)
                     /note="BLASTX similarity to P52737 (2. .45); match: 0.72,
                     score: 3.6e-13; database searched: nr; ZINC FINGER PROTEIN
                     136 pir||B57785 zinc finger protein ZNF136 - human
                     >gi|487785 (U09367) zinc finger protein ZNF136 [Homo"
     misc_feature    complement(2728. .2850)
                     /note="BLASTX similarity to Q13360 (14. .54); match: 0.7,
                     score: 2.8e-11; database searched: nr; ZINC FINGER PROTEIN
                     177 gi|1049301 (U37263) KRAB zinc finger protein, Method:
                     conceptual translation supplied by author"
     repeat_region   3149. .3444
                     /rpt_family="AluJo"
     repeat_region   3508. .3615
                     /rpt_family="FLAM_C"
     repeat_region   3691. .3826
                     /rpt_family="AluSq/x"
     repeat_region   3842. .4140
                     /rpt_family="AluSc"
     repeat_region   4343. .4643
                     /rpt_family="AluJb"
     repeat_region   4645. .4942
                     /rpt_family="AluSx"
     misc_feature    complement(5318. .5487)
                     /note="predicted exon, program: grail2exons_human_1.3,
                     frame: 2, quality: good, score: 61.000"
     repeat_region   complement(5519. .5817)
                     /rpt_family="AluSg"
     repeat_region   complement(5822. .6004)
                     /rpt_family="AluJo"
     repeat_region   complement(6267. .6530)
                     /rpt_family="AluSq"
     repeat_region   complement(7156. .7458)
                     /rpt_family="AluSx"
     misc_feature    complement(8786. .8970)
                     /note="BLASTN similarity to cpg29c4.ft1a (4. .188); match:
                     0.97, score: 1.3e-66; database searched: CpG; bases 117 to
                     304 (SL to QR)"
     repeat_region   8982. .9278
                     /rpt_family="AluSq"
     repeat_region   9290. .9513
                     /rpt_family="MER65A"
     repeat_region   9514. .9731
                     /rpt_family="AluJb"
     repeat_region   9733. .9847
                     /rpt_family="MER65A"
     repeat_region   10127. .10426
                     /rpt_family="AluSx"
     repeat_region   complement(10679. .10791)
                     /rpt_family="AluSx"
     repeat_region   11016. .11317
                     /rpt_family="AluSp"
     repeat_region   11731. .12029
                     /rpt_family="AluY"
     repeat_region   12054. .12353
                     /rpt_family="AluSg"
     repeat_region   12656. .12802
                     /rpt_family="MLT1B"
     repeat_region   complement(12798. .12956)
                     /rpt_family="MER20"
     repeat_region   complement(13091. .13231)
                     /rpt_family="MIR"
     repeat_region   14704. .14836
                     /rpt_family="MER30"
     repeat_region   complement(14844. .15142)
                     /rpt_family="AluY"
     repeat_region   15143. .15241
                     /rpt_family="MER30"
     repeat_region   15591. .15892
                     /rpt_family="AluSp"
     repeat_region   complement(15981. .16292)
                     /rpt_family="AluSx"
     repeat_region   16332. .16449
                     /rpt_family="AluSg/x"
     repeat_region   complement(16776. .16923)
                     /rpt_family="LTR8"
     repeat_region   complement(16926. .17190)
                     /rpt_family="AluSx"
     repeat_region   17372. .17677
                     /rpt_family="AluSx"
     repeat_region   17679. .18034
                     /rpt_family="MER41B"
     repeat_region   18035. .18295
                     /rpt_family="AluSp"
     repeat_region   complement(18345. .18495)
                     /rpt_family="LINE2"
     repeat_region   18510. .18817
                     /rpt_family="AluSx"
     repeat_region   18839. .18931
                     /rpt_family="MER2"
     repeat_region   18933. .19230
                     /rpt_family="AluJb"
     repeat_region   19236. .19534
                     /rpt_family="AluSq"
     repeat_region   19549. .19672
                     /rpt_family="MER2"
     repeat_region   complement(19754. .19922)
                     /rpt_family="FRAM"
     repeat_region   complement(20111. .20151)
                     /rpt_family="(CAAAA)n"
     repeat_region   complement(20152. .20459)
                     /rpt_family="AluSp"
```

```
     repeat_region   20559. .20695
                     /rpt_family="L1MC4"
     repeat_region   20700. .20866
                     /rpt_family="AluJo"
     repeat_region   complement(20867. .20979)
                     /rpt_family="AluJ"
     repeat_region   complement(20984. .21282)
                     /rpt_family="AluSg1"
     repeat_region   complement(21283. .21405)
                     /rpt_family="AluJ"
     repeat_region   21780. .22091
                     /rpt_family="AluSx"
     repeat_region   complement(22172. .22438)
                     /rpt_family="L1MD2"
     repeat_region   22612. .22801
                     /rpt_family="AluJo"
     repeat_region   22803. .22854
                     /rpt_family="(GAAA)n"
     repeat_region   22855. .22931
                     /rpt_family="(CAGG)n"
     repeat_region   22932. .23233
                     /rpt_family="AluSq"
     repeat_region   complement(23234. .23345)
                     /rpt_family="LINE2"
     repeat_region   23353. .23542
                     /rpt_family="AluJo"
     repeat_region   complement(23665. .23833)
                     /rpt_family="AluSg/x"
     repeat_region   complement(23834. .24136)
                     /rpt_family="AluSx"
     repeat_region   complement(24137. .24280)
                     /rpt_family="AluJb"
     repeat_region   24311. .24609
                     /rpt_family="AluY"
     repeat_region   24614. .24808
                     /rpt_family="MER5A"
     repeat_region   24827. .24926
                     /rpt_family="FLAM_C"
     repeat_region   24934. .25223
                     /rpt_family="AluSg"
     repeat_region   complement(25261. .25556)
                     /rpt_family="AluY"
     repeat_region   25558. .25683
                     /rpt_family="L1MA10"
     repeat_region   complement(25983. .26284)

Alignment Scores:
Pred. No.:              8.68e+03        Length:         42642
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     8               Gaps:           0

US-09-858-332G-28 (1-6) x AC006277 (1-42642)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      25600 CACAACCATAACCACAAT 25583

RESULT 437
SCCEN12RG/c
LOCUS       SCCEN12RG              43676 bp    DNA     linear   PLN 18-APR-2005
DEFINITION  S.cerevisiae DNA from CEN12 region including ATS/APSG, SCD25, SOF1,
            DRS1, MMM1 and DNM1 genes.
ACCESSION   X91488
VERSION     X91488.1  GI:1495203
KEYWORDS    ATS/APSG gene; DNM1 gene; DRS1 gene; MMM1 gene; SCD25 gene; sof1
            gene.
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 43676)
  AUTHORS   Damak,F., Boy-Marcotte,E., Le-Roscouet,D., Guilbaud,R. and
            Jacquet,M.
  TITLE     SDC25, a CDC25-like gene which contains a RAS-activating domain and
            is a dispensable gene of Saccharomyces cerevisiae
  JOURNAL   Mol. Cell. Biol. 11 (1), 202-212 (1991)
   PUBMED   1986220
REFERENCE   2  (bases 1 to 43676)
  AUTHORS   Ripmaster,T.L., Vaughn,G.P. and Woolford,J.L. Jr.
  TITLE     A putative ATP-dependent RNA helicase involved in Saccharomyces
            cerevisiae ribosome assembly
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 89 (23), 11131-11135 (1992)
   PUBMED   1454790
REFERENCE   3  (bases 1 to 43676)
  AUTHORS   Burgess,S.M., Delannoy,M. and Jensen,R.E.
  TITLE     MMM1 encodes a mitochondrial outer membrane protein essential for
            establishing and maintaining the structure of yeast mitochondria
  JOURNAL   J. Cell Biol. 126 (6), 1375-1391 (1994)
   PUBMED   8089172
REFERENCE   4  (bases 1 to 43676)
  AUTHORS   Gammie,A.E., Kurihara,L.J., Vallee,R.B. and Rose,M.D.
  TITLE     DNM1, a dynamin-related gene, participates in endosomal trafficking
            in yeast
  JOURNAL   J. Cell Biol. 130 (3), 553-566 (1995)
   PUBMED   7622557
REFERENCE   5
  AUTHORS   Miosga,T. and Zimmermann,F.K.
  TITLE     Sequence analysis of the CEN12 region of Saccharomyces cerevisiae
            on a 43.7 kb fragment of chromosome XII including an open reading
            frame homologous to the human cystic fibrosis transmembrane
            conductance regulator protein CFTR
  JOURNAL   Yeast 12 (7), 693-708 (1996)
   PUBMED   8810043
REFERENCE   6  (bases 1 to 43676)
  AUTHORS   Miosga,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-SEP-1995) T. Miosga, Institut fuer Mikrobiologie und
            Genetik, TH Darmstadt, Schnittspahnstr. 10, 64287 Darmstadt, FRG
COMMENT     Related sequences M26647, M31771, L32793, L40588, X70052, X06665,
            L00683, X03606, T37634, T36554, T36641, Z16203, Z13731, Z15219,
            Z12685.
FEATURES             Location/Qualifiers
     source          1. .43676
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /strain="FY23 /RD005 (derived from S288C)"
                     /db_xref="taxon:4932"
                     /chromosome="XII"
                     /clone_lib="cosmid clone 2B18"
     gene            <1. .423
                     /gene="L1301"
     CDS             <1. .423
                     /gene="L1301"
                     /codon_start=1
                     /product="L1301 protein"
                     /protein_id="CAA62773.1"
                     /db_xref="GI:1495205"
                     /db_xref="InterPro:IPR000583"
                     /db_xref="UniProt/TrEMBL:Q92276"
                     /translation="DQWQLTVVVLCGVFGSGFLFTGFSFSFQLESSFLLLSLFLLLFL
                     QFLQGLLGQWLTIFPKNNLSRFCGFFNSFDNIFVLRHHVKCFFFSFTINLKRPFQPMI
                     ENLNKSVVLLHPYMTLNNFSFLLFSRKCVIRYSSLMIS"
     gene            complement(<1. .243)
                     /gene="L1295/ATS/APSG"
     CDS             complement(<1. .243)
                     /gene="L1295/ATS/APSG"
                     /codon_start=1
                     /product="L1295/ATS/APSG protein"
                     /protein_id="CAA62772.1"
                     /db_xref="GI:1495204"
                     /db_xref="GOA:P04802"
                     /db_xref="UniProt/Swiss-Prot:P04802"
                     /translation="MSQDENIVKAVEESAEPAQVILGEDGKPLSKKALKKLQKEQEKQ
```

```
                RKKEERALQLEAEREAREKKAAAEDTAKDNYGKLPLI"
gene            902. .1213
                /gene="L1305/SCD25"
CDS             902. .1213
                /gene="L1305/SCD25"
                /note="an /T deletion in the sequence leads to two orfs"
                /codon_start=1
                /product="L1305/SCD25 protein"
                /protein_id="CAA62774.1"
                /db_xref="GI:1495206"
                /db_xref="GOA:P14771"
                /db_xref="InterPro:IPR001452"
                /db_xref="SGD:YLL017W"
                /db_xref="UniProt/Swiss-Prot:P14771"
                /translation="MSCTASYAGMTTPVKDKEGHGIPCLQPIDVVECTYQYFTKSRNK
                LSLRVGDLIYVLTKGSNGWWDGVLIRHSANNNNNSLILDRGWFPPSFYTVHSKRTTRG
                A"
gene            1515. .4661
                /gene="L1309/SCD25"
CDS             1515. .4661
                /gene="L1309/SCD25"
                /codon_start=1
                /product="L1309/SCD25 protein"
                /protein_id="CAA62775.1"
                /db_xref="GI:1495207"
                /db_xref="GOA:P14771"
                /db_xref="SGD:YLL016W"
                /db_xref="UniProt/Swiss-Prot:P14771"
                /translation="MPITSSPDLFYLNDCDVVYWYDLTRLVCHYVNLTERDLLANERE
                KFLTSLDLLTAQITYVYMLFRNLRLVEDSFKKTLKKLIYTLSRFSINANIWFHSTLFE
                EREAIASQKDPERRSPLLQSILGTFQKFHFLLRLLHFLSNPNELTILPQLTPRFFKDS
                FNTISWNNPFLRKRLNQHMSHDLPRQMIKAVAGASGIVAENIDEIPASKQGTSCSSET
                SHHSPSAPFQRRRRGTIFSNVSGSSDESDTIWSKRKKPYPLNEETLSLVRARKKQLDG
                KLKQMIKSANEYLSNTANFSKMLNFEMNFKTYEEVSGTIPIIDILENLDLTIFLNLRE
                LGDENRVFDEDVAIDDEDEEFLKHSLSSLSYILSDYFNMKQYFHDVVVKFIIVAQHLT
                LEDPFVFSPMQNDLPTGYYEPMKPSSLNLDNAKDKKNGSQNTDIQEEEDEYEPDPDSL
                ILFHNLINQDSDFNDLKFFNLAHVFKKSCDDYFDVLKLAIEFVNQLILERENLLNYAA
                RMMKNNITELLLRGEEGYGSYDGGETAEKSDTNAVYADSDTKDNDEWRDSQVKLPRYL
                QREYDSELIWGSNNRIKGGSKHALISYLTDNEKKDLFFDITFLITFRSIFTTTEFLSY
                LISQYNLDPPEDLCFEEYNEWVTKKLIPVKCRVVEIMTTFFKQYWFLGYDEPDLATLN
                LDYFAQVAIKENITGSVELLKEVNQKFKHGNIQEATAPMKTLDQQICQDHYSGTLYST
                TESILAVDPVLFATQLTILEHEIYCEITIFDCLQKIWKNKYTKSYGASPGLNEFISFA
                NKLTNFISYSVVKEADKSKRAKLLSHFIFIAEYCRKFNNFSSMTAIISALYSSPIYRL
                EKTWQAVIPQTRDLLQSLNKLMDPKKNFINYRNELKSLHSAPCVPFFGVYLSDLTFTD
                SGNPDYLVLEHGLKGVHDEKKYINFNKRSRLVDILQEIIYFKKTHYDFTKDRTVIECI
                SNSLENIPHIEKQYQLSLIIEPKPRKKVVPNSNSNNKSQEKSRDDQTDEGKTSTKKDR
                FPKFQLHKTKKKAPKVSK"
conflict        1515. .4661
                /gene="L1309/SCD25"
                /citation=[1]
gene            5100. .9779
                /gene="L1313"
CDS             5100. .9779
                /gene="L1313"
                /codon_start=1
                /product="L1313 protein"
                /protein_id="CAA62776.1"
                /db_xref="GI:1495208"
                /db_xref="GOA:P14772"
                /db_xref="InterPro:IPR001140"
                /db_xref="InterPro:IPR003439"
                /db_xref="InterPro:IPR003593"
                /db_xref="SGD:S000003938"
                /db_xref="UniProt/Swiss-Prot:P14772"
                /translation="MSSLEVVDGCPYGYRPYPDSGTNALNPCFISVISAWQAVFFLLI
                GSYQLWKLYKNNKVPPRFKNFPTLPSKINSRHLTHLTNVCFQSTLIICELALVSQSSD
                RVYPFILKKALYLNLLFNLGISLPTQYLAYFKSTFSMGNQLFYYMFQILLQLFLILQR
                YYHGSSNERLTVISGQTAMILEVLLLFNSVAIFIYDLCIFEPINELSEYYKKNGWYPP
                VHVLSYITFIWMNKLIVETYRNKKIKDPNQLPLPPVDLNIKSISKEFKANWELEKWLN
                RNSLWRAIWKSFGRTISVAMLYETTSDLLSVVQPQFLRIFIDGLNPETSSKYPPLNGV
                FIALTLFVISVVSVFLTNQFYIGIFEAGLGIRGSLASLVYQKSLRLTLAERNEKSTGD
                ILNLMSVDVLRIQRFFENAQTIIGAPIQIIVVLTSLYWLLGKAVIGGLVTMAIMMPIN
                AFLSRKVKKLSKTQMKYKDMRIKTITELLNAIKSIKLYAWEEPMMARLNHVRNDMELK
                NFRKIGIVSNLIYFAWNCVPLMVTCSTFGLFSLFSDSPLSPAIVFPSLSLFNILNSAI
                YSVPSMINTIIETSVSMERLKSFLLSDEIDDSFIERIDPSADERALPAIEMNNITFLW
                KSKEVLTSSQSGDNLRTDEESIIGSSQIALKNIDHFEAKRGDLVCVVGRVGAGKSTFL
                KAILGQLPCMSGSRDSIPPKLIIRSSSVAYCSQESWIMNASVRENILFGHKFDQDYYD
                LTIKACQLLPDLKILPDGDETLVGEKGISLSGGQKARLSLARAVYSRADIYLLDDILS
                AVDAEVSKNIIEYVLIGKTALLKNKTIILTTNTVSILKHSQMIYALENGEIVEQGNYE
                DVMNRKNNTSKLKKLLEEFDSPIDNGNESDVQTEHRSESEVDEPLQLKVTESETEDEV
                VTESELELIKANSRRASLATLRPRPFVGAQLDSVKKTAQKAEKTEVGRVKTKIYLAYI
                KACGVLGVVLFFLFMILTRVFDLAENFWLKYWSESNEKNGSNERVWMFVGVYSLIGVA
                SAAFNNLRSIMMLLYCSIRGSKKLHESMAKSVIRSPMTFFETTPVGRIINRFSSDMDA
                VDSNLQYIFSFFFKSILTYLVTVILVGYNMPWFLVFNMFLVVIYIYYQTFYIVLSREL
                KRLISISYSPIMSLMSESLNGYSIIDAYDHFERFIYLNYEKIQYNVDFVFNFRSTNRW
                LSVRLQTIGATIVLATAILALATMNTKRQLSSGMVGLLMSYSLEVTGSLTWIVRTTVT
                IETNIVSVERIVEYCELPPEAQSINPEKRPDENWPSKGGIEFKNYSTKYRENLDPVLN
                NINVKIEPCEKVGIVGRTGAGKSTLSLALFRILEPTEGKIIIDGIDISDIGLFDLRSH
                LAIIPQDAQAFEGTVKTNLDPFNRYSEDELKRAVEQAHLKPHLEKMLHSKPRGDDSNE
                EDGNVNDILDVKINENGSNLSVGQRQLLCLARALLNRSKILVLDEATASVDMETDKII
                QDTIRREFKDRTILTIAHRIDTVLDSDKIIVLDQGSVREFDSPSKLLSDKTSIFYSLC
                EKGGYLK"
gene            complement(8000. .8386)
                /gene="L1317 (internal to L1313)"
CDS             complement(8000. .8386)
                /gene="L1317 (internal to L1313)"
                /codon_start=1
                /product="L1317 protein"
                /protein_id="CAA62777.1"
                /db_xref="GI:1495209"
                /db_xref="UniProt/TrEMBL:Q92277"
                /translation="MYCRLLSTASISDENLFMILPTGVVSKKVIGLLITDLAMLSCSF
                LEPLIEQYSSIIILRKLLNAADATPIKEYTPTNIQTLSFEPFFSLDSDQYFNQKFSAK
                SKTLVNIINKKNKTTPKTPQALIYAR"
gene            9990. .10316
                /gene="L1321"
CDS             9990. .10316
                /gene="L1321"
                /codon_start=1
                /product="L1321 protein"
                /protein_id="CAA62778.1"
                /db_xref="GI:1495210"
                /db_xref="GOA:Q12431"
                /db_xref="SGD:S000003937"
                /db_xref="UniProt/TrEMBL:Q12431"
                /translation="MSSNEEVFTQINATANVVDNKKRLLFVQDSSALVLGLVAGFLQI
```

```
Alignment Scores:
Pred. No.:              8.88e+03        Length:         43676
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x SCCEN12RG (1-43676)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      18794 CACAACCACAACCACAAC 18777

RESULT 438
FR151J19/c
LOCUS       FR151J19               45565 bp    DNA     linear   VRT 16-APR-2005
DEFINITION  Fugu rubripes cosmid 151J19 covering the WT1, reticulocalbin and
            PAX6 genes.
ACCESSION   AL021531
VERSION     AL021531.1  GI:3402198
KEYWORDS    .
SOURCE      Takifugu rubripes (Fugu rubripes)
  ORGANISM  Takifugu rubripes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Takifugu.
REFERENCE   1
```

```
  AUTHORS   Miles,C., Elgar,G., Coles,E., Kleinjan,D.J., van Heyningen,V. and
            Hastie,N.
  TITLE     Complete sequencing of the Fugu WAGR region from WT1 to PAX6:
            dramatic compaction and conservation of synteny with human
            chromosome 11p13
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (22), 13068-13072 (1998)
   PUBMED   9789042
REFERENCE   2  (bases 1 to 45565)
  AUTHORS   Elgar,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JAN-1998) Elgar G., UK HGMP Resource Centre, Wellcome
            Trust Genome Campus, Hinxton Hall, Hinxton, Cambs. CB10 1SB UK
FEATURES             Location/Qualifiers
     source          1. .45565
                     /organism="Takifugu rubripes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:31033"
     gene            <191. .>5375
                     /gene="WT1"
     mRNA            join(<191. .578,2117. .2239,2702. .2801,3714. .3794,
                     4011. .4116,4201. .4348,4434. .4523,5019. .5111,5254. .>5375)
                     /gene="WT1"
     CDS             join(191. .578,2117. .2239,2702. .2801,3714. .3794,4011. .4116,
                     4201. .4348,4434. .4523,5019. .5111,5254. .5375)
                     /gene="WT1"
                     /note="Exon 8 has an alternate splice site at base 5103
                     generating an exon 9bp shorter and removing the amino acid
                     sequence KPS"
                     /codon_start=1
                     /product="Wilms tumour gene"
                     /protein_id="CAA16491.1"
                     /db_xref="GI:3402199"
                     /db_xref="GOA:O93433"
                     /db_xref="InterPro:IPR000976"
                     /db_xref="InterPro:IPR007087"
                     /db_xref="UniProt/TrEMBL:O93433"
                     /translation="MGTDVRDLTLLPPAPPVSSIPGAGSGCGMSVGGGQWTQLLDLHP
                     SSPYSTLSSHHSLIKQEPSWGSSDPVDDPHCGLGAFTLHFSGQFTGPGPCRVGAFGEP
                     PAGQPRMFPNGTCLPSCMDSPPPPRNQGYGAASLDNNPNYGHTPSHHPPQFSSLPFKH
                     EDTLSPPNNIVEQQFPAPSAMFGCHNPPESCPSSQALLLRNYSSDNLYQMASQLDCVT
                     WNQMNTLASSMKSGHTSSFETEPIAPPPPVLVSAQYHIHAHSVFRGIQDVCRVSSIAP
                     PVVKSSDTTEKRPFVCAYPGCSKKYFKLSHLQMHGRKHTGEKPYQCDVTECGRRFSRS
                     DQLKRHQLRHTGVKSFECETCQRRFARSDHLKTHTRTHTGKPSEKPFICRWSNCQKKF
                     ARSDELTRHHSMHQRNLTKLQPAI"
     exon            <191. .578
                     /gene="WT1"
                     /number=1
     intron          579. .2116
                     /gene="WT1"
                     /number=1
     exon            2117. .2239
                     /gene="WT1"
                     /number=2
     intron          2240. .2701
                     /gene="WT1"
                     /number=2
     exon            2702. .2801
                     /gene="WT1"
                     /number=3
     intron          2802. .3713
                     /gene="WT1"
                     /number=3
     repeat_region   3226. .3273
                     /note="CA repeat"
                     /rpt_type=TANDEM
     repeat_region   3510. .3564
                     /note="GT repeat"
                     /rpt_type=TANDEM
     exon            3714. .3794
                     /gene="WT1"
                     /number=4
     intron          3795. .4010
                     /gene="WT1"
                     /number=4
     exon            4011. .4116
                     /gene="WT1"
                     /number=5
     intron          4117. .4200
                     /gene="WT1"
                     /number=5
     exon            4201. .4348
                     /gene="WT1"
                     /number=6
     intron          4349. .4433
                     /gene="WT1"
                     /number=6
     exon            4434. .4523
                     /gene="WT1"
                     /number=7
     intron          4524. .5018
                     /gene="WT1"
                     /number=7
     exon            5019. .5111
                     /gene="WT1"
                     /number=8
     intron          5112. .5253
                     /gene="WT1"
                     /number=8
     exon            5254. .>5375
                     /gene="WT1"
                     /number=9
     repeat_region   5381. .5413
                     /note="AGG repeat"
                     /rpt_type=TANDEM
     repeat_region   5542. .5580
                     /note="GT repeat"
                     /rpt_type=TANDEM
     repeat_region   6704. .6735
                     /note="GT repeat"
                     /rpt_type=TANDEM
     repeat_region   7168. .7195
                     /note="ATC repeat"
                     /rpt_type=TANDEM
     repeat_region   7774. .7795
                     /note="GT repeat"
                     /rpt_type=TANDEM
     repeat_region   12413. .12512
                     /note="CA repeat"
                     /rpt_type=TANDEM
     gene            complement(<12515. .>16300)
                     /gene="RCN1"
     mRNA            complement(join(<12515. .12622,12702. .12901,13148. .13208,
                     14439. .14617,14685. .14878,16074. .>16300))
                     /gene="RCN1"
     CDS             complement(join(12515. .12622,12702. .12901,13148. .13208,
                     14439. .14617,14685. .14878,16074. .16300))
                     /gene="RCN1"
                     /codon_start=1
                     /product="Reticulocalbin"
                     /protein_id="CAA16492.1"
                     /db_xref="GI:3402200"
                     /db_xref="GOA:O93434"
                     /db_xref="InterPro:IPR000886"
                     /db_xref="InterPro:IPR002048"
                     /db_xref="UniProt/Swiss-Prot:O93434"
                     /translation="MDVLGFICAVFLGTVVLHAKPTVRKDRVLYQDPELLRQATNEDN
                     NSFQYDHEVFLGKEESKTFDQLSPEESKDRLSKIVDRIDGDGNSYITTDELKAWIKRV
                     QKRYVYENVVKVWADYDLNKDNKISWEEYKQATYGYYLSNPEEFDETTDQFSFKKMLP
                     RDERRFKRADLDGDSAANREEFTSFLHPEEFEHMKDIVVLETLEDIDKNSDGHVDEDE
                     YIADMFAHEDRGPEPEWVKTEREQFSDFRDLNKDGKMDLDEIRHWIMPQDYDHAQAEA
                     RHLVYESDKDKDQMLTKEEILDNWNMFVGSQATNYGEDLTRNHDEL"
     exon            complement(<12515. .12622)
                     /gene="RCN1"
                     /number=6
     intron          complement(12621. .12703)
                     /gene="RCN1"
```

```
                     /number=5
     exon            complement(12702. .12901)
                     /gene="RCN1"
                     /number=5
     intron          complement(12900. .13149)
                     /gene="RCN1"
                     /number=4
     exon            complement(13148. .13208)
                     /gene="RCN1"
                     /number=4
     intron          complement(13207. .14440)
                     /gene="RCN1"
                     /number=3
     repeat_region   13595. .13680
                     /note="ATCC repeat"
                     /rpt_type=TANDEM
     exon            complement(14439. .14617)
                     /gene="RCN1"
                     /number=3
     intron          complement(14616. .14686)
                     /gene="RCN1"
                     /number=2
     exon            complement(14685. .14878)
                     /gene="RCN1"
                     /number=2
     intron          complement(14877. .16075)
                     /gene="RCN1"
                     /number=1
     repeat_region   15020. .15100
                     /note="GT rich minisatellite"
                     /rpt_family="minisatellite"
                     /rpt_type=TANDEM
                     /rpt_unit="15020. .15052"
     exon            complement(16074. .>16300)
                     /gene="RCN1"
                     /number=1
     repeat_region   19872. .19940
                     /note="GT repeat"
                     /rpt_type=TANDEM
     repeat_region   20129. .20176
                     /note="GCT repeat"

Alignment Scores:
Pred. No.:                9.25e+03        Length:         45565
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       5               Gaps:           0

US-09-858-332G-28 (1-6) x FR151J19 (1-45565)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      15417 CATAACCATAACCATAAC 15400

RESULT 439
AC136498/c
LOCUS       AC136498               46165 bp    DNA     linear   HTG 06-NOV-2002
DEFINITION  Drosophila pseudoobscura clone DPSF1-548A4, *** SEQUENCING IN
            PROGRESS ***.
ACCESSION   AC136498
VERSION     AC136498.1  GI:24635351
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila pseudoobscura
  ORGANISM  Drosophila pseudoobscura
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 46165)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Adams,C., Allen,C., Allen,H.,
            Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Baca,E.,
            Bandaranaike,D., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P.,
            Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Chacko,J.,
            Chavez,D., Chen,G., Chen,R., Chen,Z., Chu,J., Cleveland,C.,
            Cockrell,R., Cox,C., Coyle,M., Cree,A., Davila,M.L., Davis,C.,
            Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S.,
            Deramo,C., Ding,Y., Dinh,H., Divya Khurana, Draper,H., D'Souza,L.,
            Dugan-Rocha,S., Dunn,A., Durbin,K., Eaves,K., Egan,A., Escotto,M.,
            Eugene,C., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N.,
            Forbes,L., Foster,M., Foster,P., Gabisi,A., Ganta,R., Garcia,A.,
            Garner,T., Garza,M., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamilton,C., Hamilton,K., Harvey,Y.,
            Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R.,
            Hines,S., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S.,
            Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H.,
            Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S.,
            Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Lebow,H., Levan,J.,
            Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P.,
            Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Malloy,K., Mangum,A., Mangum,B.,
            Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S.,
            Mcleod,M., Mcneill,T., Milosavljevic,A., Miner,G., Minja,E.,
            Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S.,
            Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N.,
            Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G.,
            Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H.,
            Perez,A., Perez,L., Plopper,F., Poindexter,A., Popovic,D.,
            Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K.,
            Reigh,R., Ren,Y., Reuter,M., Richards,S., Rives,C., Rodkey,T.,
            Rojas,A., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G.,
            Shen,H., Sisson,I., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z.,
            Sorelle,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T.,
            Thomas,N., Thomas,S., Trejos,Z., Usmani,K., Valas,R., Vera,V.,
            Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S.,
            Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R.,
            Wleczyk,R., Wooden,H., Worley,K., Wright,R., Wu,J., Yakub,S.,
            Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 46165)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     ---------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            ---------- Project Information
                Center project name: OBAD
                Center clone name: DPSF1-548A4
            ---------- Summary Statistics
                Sequencing vector: Plasmid;
                Sequencing vector: M13;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 48837 bases at least Q40
                Consensus quality: 49308 bases at least Q30
                Consensus quality: 49614 bases at least Q20
            ----------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
```

```
          *        1    46165: contig of 46165 bp in length.
FEATURES           Location/Qualifiers
     source        1. .46165
                   /organism="Drosophila pseudoobscura"
                   /mol_type="genomic DNA"
                   /db_xref="taxon:7237"
                   /clone="DPSF1-548A4"
ORIGIN

Alignment Scores:
Pred. No.:              9.37e+03       Length:        46165
Score:                  42.00          Matches:       6
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     14             Gaps:          0

US-09-858-332G-28 (1-6) x AC136498 (1-46165)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        359 CATAACCATAACCATAAT 342

RESULT 440
AC015044
LOCUS       AC015044               47919 bp    DNA     linear   HTG 16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC015044
VERSION     AC015044.1  GI:6436291
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 47919)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10211036 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES           Location/Qualifiers
     source        1. .47919
                   /organism="Drosophila melanogaster"
                   /mol_type="genomic DNA"
                   /db_xref="taxon:7227"
ORIGIN

Alignment Scores:
Pred. No.:              9.71e+03       Length:        47919
Score:                  42.00          Matches:       6
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     14             Gaps:          0

US-09-858-332G-28 (1-6) x AC015044 (1-47919)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       3280 CACAATCACAATCACAAT 3297

RESULT 441
AC132900/c
LOCUS       AC132900               49628 bp    DNA     linear   HTG 04-SEP-2002
DEFINITION  Homo sapiens chromosome 17 clone RP11-1427K10 map 17, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC132900
VERSION     AC132900.1  GI:22711714
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 49628)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-1427K10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 49628)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-SEP-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L28156
                Center clone name: 1427_K_10
            --------------.
            * NOTE: This record contains 62 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      695: contig of 695 bp in length
            *      696      795: gap of 100 bp
            *      796     1491: contig of 696 bp in length
            *     1492     1591: gap of 100 bp
            *     1592     2290: contig of 699 bp in length
            *     2291     2390: gap of 100 bp
            *     2391     3095: contig of 705 bp in length
            *     3096     3195: gap of 100 bp
            *     3196     3899: contig of 704 bp in length
            *     3900     3999: gap of 100 bp
            *     4000     4701: contig of 702 bp in length
            *     4702     4801: gap of 100 bp
            *     4802     5487: contig of 686 bp in length
            *     5488     5587: gap of 100 bp
            *     5588     6304: contig of 717 bp in length
            *     6305     6404: gap of 100 bp
            *     6405     7113: contig of 709 bp in length
```

```
*     7114     7213: gap of 100 bp
*     7214     7933: contig of 720 bp in length
*     7934     8033: gap of 100 bp
*     8034     8725: contig of 692 bp in length
*     8726     8825: gap of 100 bp
*     8826     9530: contig of 705 bp in length
*     9531     9630: gap of 100 bp
*     9631    10349: contig of 719 bp in length
*    10350    10449: gap of 100 bp
*    10450    11148: contig of 699 bp in length
*    11149    11248: gap of 100 bp
*    11249    11947: contig of 699 bp in length
*    11948    12047: gap of 100 bp
*    12048    12763: contig of 716 bp in length
*    12764    12863: gap of 100 bp
*    12864    13567: contig of 704 bp in length
*    13568    13667: gap of 100 bp
*    13668    14360: contig of 693 bp in length
*    14361    14460: gap of 100 bp
*    14461    15152: contig of 692 bp in length
*    15153    15252: gap of 100 bp
*    15253    15959: contig of 707 bp in length
*    15960    16059: gap of 100 bp
*    16060    16755: contig of 696 bp in length
*    16756    16855: gap of 100 bp
*    16856    17552: contig of 697 bp in length
*    17553    17652: gap of 100 bp
*    17653    18351: contig of 699 bp in length
*    18352    18451: gap of 100 bp
*    18452    19158: contig of 707 bp in length
*    19159    19258: gap of 100 bp
*    19259    19951: contig of 693 bp in length
*    19952    20051: gap of 100 bp
*    20052    20762: contig of 711 bp in length
*    20763    20862: gap of 100 bp
*    20863    21564: contig of 702 bp in length
*    21565    21664: gap of 100 bp
*    21665    22381: contig of 717 bp in length
*    22382    22481: gap of 100 bp
*    22482    23175: contig of 694 bp in length
*    23176    23275: gap of 100 bp
*    23276    23983: contig of 708 bp in length
*    23984    24083: gap of 100 bp
*    24084    24770: contig of 687 bp in length
*    24771    24870: gap of 100 bp
*    24871    25568: contig of 698 bp in length
*    25569    25668: gap of 100 bp
*    25669    26380: contig of 712 bp in length
*    26381    26480: gap of 100 bp
*    26481    27170: contig of 690 bp in length
*    27171    27270: gap of 100 bp
*    27271    27972: contig of 702 bp in length
*    27973    28072: gap of 100 bp
*    28073    28785: contig of 713 bp in length
*    28786    28885: gap of 100 bp
*    28886    29603: contig of 718 bp in length
*    29604    29703: gap of 100 bp
*    29704    30414: contig of 711 bp in length
*    30415    30514: gap of 100 bp
*    30515    31220: contig of 706 bp in length
*    31221    31320: gap of 100 bp
*    31321    31979: contig of 659 bp in length
*    31980    32079: gap of 100 bp
*    32080    32783: contig of 704 bp in length
*    32784    32883: gap of 100 bp
*    32884    33589: contig of 706 bp in length
*    33590    33689: gap of 100 bp
*    33690    34392: contig of 703 bp in length
*    34393    34492: gap of 100 bp
*    34493    35191: contig of 699 bp in length
*    35192    35291: gap of 100 bp
*    35292    35983: contig of 692 bp in length
*    35984    36083: gap of 100 bp
*    36084    36788: contig of 705 bp in length
*    36789    36888: gap of 100 bp
*    36889    37606: contig of 718 bp in length
*    37607    37706: gap of 100 bp
*    37707    38402: contig of 696 bp in length
*    38403    38502: gap of 100 bp
*    38503    39208: contig of 706 bp in length
*    39209    39308: gap of 100 bp
*    39309    40003: contig of 695 bp in length
*    40004    40103: gap of 100 bp
*    40104    40806: contig of 703 bp in length
*    40807    40906: gap of 100 bp
*    40907    41598: contig of 692 bp in length
*    41599    41698: gap of 100 bp
*    41699    42401: contig of 703 bp in length
*    42402    42501: gap of 100 bp
*    42502    43204: contig of 703 bp in length
*    43205    43304: gap of 100 bp
*    43305    44008: contig of 704 bp in length
*    44009    44108: gap of 100 bp
*    44109    44811: contig of 703 bp in length
*    44812    44911: gap of 100 bp
*    44912    45620: contig of 709 bp in length
*    45621    45720: gap of 100 bp
*    45721    46419: contig of 699 bp in length
*    46420    46519: gap of 100 bp
*    46520    47226: contig of 707 bp in length
*    47227    47326: gap of 100 bp
*    47327    48037: contig of 711 bp in length
*    48038    48137: gap of 100 bp
*    48138    48839: contig of 702 bp in length
*    48840    48939: gap of 100 bp
*    48940    49628: contig of 689 bp in length.
FEATURES             Location/Qualifiers
     source          1. .49628
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-1427K10"
                     /clone_lib="RPCI-11 Male BAC segment 5"
     gap             696. .795
                     /estimated_length=100
     gap             1492. .1591
                     /estimated_length=100
     gap             2291. .2390
                     /estimated_length=100
     gap             3096. .3195
                     /estimated_length=100

Alignment Scores:
Pred. No.:                1e+04          Length:          49628
Score:                    42.00          Matches:         6
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       14             Gaps:            0

US-09-858-332G-28 (1-6) x AC132900 (1-49628)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      13419 CACAATCATAACCACAAC 13402

RESULT 442
AC090162/c
LOCUS       AC090162               49688 bp    DNA     linear   HTG 21-DEC-2001
DEFINITION  Homo sapiens chromosome 11 clone RP11-684M15 map 11, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC090162
VERSION     AC090162.4  GI:17975285
KEYWORDS    HTG; HTGS_PHASE0.
```

```
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 49688)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-684M15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 49688)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Dec 21, 2001 this sequence version replaced gi:14141771.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L11855
                Center clone name: 684_M_15
            --------------.
            * NOTE: This record contains 60 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      709: contig of 709 bp in length
            *      710      809: gap of 100 bp
            *      810     1545: contig of 736 bp in length
            *     1546     1645: gap of 100 bp
            *     1646     2403: contig of 758 bp in length
            *     2404     2503: gap of 100 bp
            *     2504     3211: contig of 708 bp in length
            *     3212     3311: gap of 100 bp
            *     3312     4040: contig of 729 bp in length
            *     4041     4140: gap of 100 bp
            *     4141     4876: contig of 736 bp in length
            *     4877     4976: gap of 100 bp
            *     4977     5724: contig of 748 bp in length
            *     5725     5824: gap of 100 bp
            *     5825     6564: contig of 740 bp in length
            *     6565     6664: gap of 100 bp
            *     6665     7433: contig of 769 bp in length
            *     7434     7533: gap of 100 bp
            *     7534     8252: contig of 719 bp in length
            *     8253     8352: gap of 100 bp
            *     8353     9100: contig of 748 bp in length
            *     9101     9200: gap of 100 bp
            *     9201     9956: contig of 756 bp in length
            *     9957    10056: gap of 100 bp
            *    10057    10782: contig of 726 bp in length
            *    10783    10882: gap of 100 bp
            *    10883    11590: contig of 708 bp in length
            *    11591    11690: gap of 100 bp
            *    11691    12382: contig of 692 bp in length
            *    12383    12482: gap of 100 bp
            *    12483    13159: contig of 677 bp in length
            *    13160    13259: gap of 100 bp
            *    13260    14014: contig of 755 bp in length
            *    14015    14114: gap of 100 bp
            *    14115    14825: contig of 711 bp in length
            *    14826    14925: gap of 100 bp
            *    14926    15615: contig of 690 bp in length
            *    15616    15715: gap of 100 bp
            *    15716    16419: contig of 704 bp in length
            *    16420    16519: gap of 100 bp
            *    16520    17262: contig of 743 bp in length
            *    17263    17362: gap of 100 bp
            *    17363    18089: contig of 727 bp in length
            *    18090    18189: gap of 100 bp
            *    18190    18952: contig of 763 bp in length
            *    18953    19052: gap of 100 bp
            *    19053    19794: contig of 742 bp in length
            *    19795    19894: gap of 100 bp
            *    19895    20629: contig of 735 bp in length
            *    20630    20729: gap of 100 bp
            *    20730    21485: contig of 756 bp in length
            *    21486    21585: gap of 100 bp
            *    21586    22314: contig of 729 bp in length
            *    22315    22414: gap of 100 bp
            *    22415    23107: contig of 693 bp in length
            *    23108    23207: gap of 100 bp
            *    23208    23911: contig of 704 bp in length
            *    23912    24011: gap of 100 bp
            *    24012    24735: contig of 724 bp in length
            *    24736    24835: gap of 100 bp
            *    24836    25587: contig of 752 bp in length
            *    25588    25687: gap of 100 bp
            *    25688    26451: contig of 764 bp in length
            *    26452    26551: gap of 100 bp
            *    26552    27247: contig of 696 bp in length
            *    27248    27347: gap of 100 bp
            *    27348    28103: contig of 756 bp in length
            *    28104    28203: gap of 100 bp
            *    28204    28901: contig of 698 bp in length
            *    28902    29001: gap of 100 bp
            *    29002    29712: contig of 711 bp in length
            *    29713    29812: gap of 100 bp
            *    29813    30504: contig of 692 bp in length
            *    30505    30604: gap of 100 bp
            *    30605    31364: contig of 760 bp in length
            *    31365    31464: gap of 100 bp
            *    31465    32229: contig of 765 bp in length
            *    32230    32329: gap of 100 bp
            *    32330    33078: contig of 749 bp in length
            *    33079    33178: gap of 100 bp
            *    33179    33934: contig of 756 bp in length
            *    33935    34034: gap of 100 bp
            *    34035    34779: contig of 745 bp in length
            *    34780    34879: gap of 100 bp
            *    34880    35591: contig of 712 bp in length
            *    35592    35691: gap of 100 bp
            *    35692    36442: contig of 751 bp in length
            *    36443    36542: gap of 100 bp
```

```
            *    36543    37285: contig of 743 bp in length
            *    37286    37385: gap of 100 bp
            *    37386    38126: contig of 741 bp in length
            *    38127    38226: gap of 100 bp
            *    38227    38944: contig of 718 bp in length
            *    38945    39044: gap of 100 bp
            *    39045    39791: contig of 747 bp in length
            *    39792    39891: gap of 100 bp
            *    39892    40605: contig of 714 bp in length
            *    40606    40705: gap of 100 bp
            *    40706    41400: contig of 695 bp in length
            *    41401    41500: gap of 100 bp
            *    41501    42222: contig of 722 bp in length
            *    42223    42322: gap of 100 bp
            *    42323    43066: contig of 744 bp in length
            *    43067    43166: gap of 100 bp
            *    43167    43917: contig of 751 bp in length
            *    43918    44017: gap of 100 bp
            *    44018    44751: contig of 734 bp in length
            *    44752    44851: gap of 100 bp
            *    44852    45559: contig of 708 bp in length
            *    45560    45659: gap of 100 bp
            *    45660    46428: contig of 769 bp in length
            *    46429    46528: gap of 100 bp
            *    46529    47204: contig of 676 bp in length
            *    47205    47304: gap of 100 bp
            *    47305    48009: contig of 705 bp in length
            *    48010    48109: gap of 100 bp
            *    48110    48870: contig of 761 bp in length
            *    48871    48970: gap of 100 bp
            *    48971    49688: contig of 718 bp in length.
FEATURES             Location/Qualifiers
     source          1. .49688
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-684M15"
                     /clone_lib="RPCI-11 Human Male BAC"
     gap             710. .809
                     /estimated_length=100
     gap             1546. .1645
                     /estimated_length=100
     gap             2404. .2503
                     /estimated_length=100
     gap             3212. .3311
                     /estimated_length=100

Alignment Scores:
Pred. No.:              1.01e+04        Length:         49688
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AC090162 (1-49688)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      36100 CATAACCATAACCATAAT 36083

RESULT 443
AC096992
LOCUS       AC096992               50966 bp    DNA     linear   PRI 02-OCT-2002
DEFINITION  Homo sapiens 3 BAC RP11-731C17 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC096992
VERSION     AC096992.15  GI:23396182
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 50966)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
            Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N.,
            Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N.,
            Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,
            Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M.,
            Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S.,
            Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E.,
            Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,
            Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,
            Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,
            Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S.,
            Warren,R., Washington,C., Watlington,S., Williams,G.,
            Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,
            Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and
            Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 50966)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 50966)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 50966)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   5  (bases 1 to 50966)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Oct 1, 2002 this sequence version replaced gi:23264923.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
```

gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:
http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

```
FEATURES             Location/Qualifiers
     source          1. .50966
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-731C17"
     misc_feature    58. .2000
                     /note="overlaps bases 188127. .190069 of clone AC011597"
                     /function="clone overlap"
     repeat_region   888. .930
                     /rpt_family="(TA)n"
     repeat_region   1036. .1327
                     /rpt_family="AluJb"
     repeat_region   1380. .1576
                     /rpt_family="L1PA9"
     repeat_region   2007. .2038
                     /rpt_family="(TG)n"
     repeat_region   2829. .3023
                     /rpt_family="L1MC4"
     repeat_region   3024. .3048
                     /rpt_family="(TG)n"
     repeat_region   3049. .4360
                     /rpt_family="L1MC4"
     repeat_region   4367. .4483
                     /rpt_family="MIR"
     repeat_region   4487. .4524
                     /rpt_family="(ATGGTG)n"
     repeat_region   4545. .4840
                     /rpt_family="AluSq"
     repeat_region   5128. .5163
                     /rpt_family="AT_rich"
     repeat_region   5239. .5268
                     /rpt_family="AT_rich"
     repeat_region   5362. .5715
                     /rpt_family="MER1B"
     repeat_region   complement(6074. .6384)
                     /rpt_family="AluY"
     repeat_region   complement(6512. .6661)
                     /rpt_family="L2"
     repeat_region   6858. .7150
                     /rpt_family="AluY"
     repeat_region   complement(7292. .7465)
                     /rpt_family="MIR"
     repeat_region   7618. .7799
                     /rpt_family="L2"
     repeat_region   8298. .8457
                     /rpt_family="L1MA2"
     repeat_region   8458. .8588
                     /rpt_family="AluSc"
     repeat_region   8589. .8895
                     /rpt_family="AluY"
     repeat_region   8896. .9070
                     /rpt_family="AluSc"
     repeat_region   9071. .9653
                     /rpt_family="L1MA2"
     repeat_region   9654. .9942
                     /rpt_family="AluSc"
     repeat_region   9943. .11044
                     /rpt_family="L1MA2"
     repeat_region   11045. .11337
                     /rpt_family="AluSx"
     repeat_region   11338. .11687
                     /rpt_family="L1MA2"
     repeat_region   11688. .11957
                     /rpt_family="AluJo"
     repeat_region   11958. .12105
                     /rpt_family="L1MA2"
     repeat_region   complement(12106. .12201)
                     /rpt_family="SVA"
     repeat_region   12202. .12237
                     /rpt_family="L1MA2"
     repeat_region   12238. .12538
                     /rpt_family="AluJb"
     repeat_region   12539. .12841
                     /rpt_family="L1MA2"
     repeat_region   12842. .13118
                     /rpt_family="AluY"
     repeat_region   13119. .13141
                     /rpt_family="(TAAAA)n"
     repeat_region   13152. .13629
                     /rpt_family="L1MA2"
     repeat_region   13630. .13929

Alignment Scores:
Pred. No.:              1.03e+04        Length:         50966
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     8               Gaps:           0

US-09-858-332G-28 (1-6) x AC096992 (1-50966)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      28862 CATAACCATAACCATAAT 28879

RESULT 444
AC019512/c
LOCUS       AC019512               51405 bp    DNA     linear   HTG 03-JAN-2000
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC019512
VERSION     AC019512.1  GI:6665385
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
```

```
REFERENCE   1  (bases 1 to 51405)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10210037 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .51405
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
ORIGIN

Alignment Scores:
Pred. No.:                1.04e+04        Length:         51405
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       14              Gaps:           0

US-09-858-332G-28 (1-6) x AC019512 (1-51405)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      50578 CATAACCATAACCATAAT 50561

RESULT 445
AL845523
LOCUS       AL845523               51413 bp    DNA     linear   ROD 24-DEC-2002
DEFINITION  Mouse DNA sequence from clone RP23-290B1 on chromosome 2, complete
            sequence.
ACCESSION   AL845523
VERSION     AL845523.6  GI:27374592
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 51413)
  AUTHORS   Tracey,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-DEC-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Dec 26, 2002 this sequence version replaced gi:23573346.
            Sequence from the Mouse Genome Sequencing Consortium whole genome
            shotgun may have been used to confirm this sequence.  Sequence data
            from the whole genome shotgun alone has only been used where it has
            a phred quality of at least 30.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-290B1 is
            from the RPCI-23 Mouse PAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .51413
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="2"
                     /clone="RP23-290B1"
                     /clone_lib="RPCI-23"
ORIGIN

Alignment Scores:
Pred. No.:                1.04e+04        Length:         51413
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       9               Gaps:           0

US-09-858-332G-28 (1-6) x AL845523 (1-51413)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      11761 CACAACCACAACCACAAT 11778

RESULT 446
AL160071
LOCUS       AL160071               51556 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP4-803K15 on chromosome 20 Contains
            the 3' end of the C20orf179 gene for chromosome 20 open reading
            frame 179, complete sequence.
ACCESSION   AL160071
VERSION     AL160071.16  GI:10443395
KEYWORDS    HTG; BAF; BCRG1; C20orf179.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 51556)
  AUTHORS   Kay,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 1, 2000 this sequence version replaced gi:10185499.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 20, constructed by the Sanger Centre Chromosome 20
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr20
            RP4-803K15 is from the library RPCI-4 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
```

```
          Web site: http://www.sanger.ac.uk
          Contact: vega@sanger.ac.uk
          --------------
          This sequence was finished as follows unless otherwise noted: all
          regions were either double-stranded or sequenced with an alternate
          chemistry or covered by high quality data (i.e., phred quality >=
          30); an attempt was made to resolve all sequencing problems, such
          as compressions and repeats; all regions were covered by at least
          one subclone; and the assembly was confirmed by restriction digest,
          except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1. .51556
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="RZPD:RPCIP704K15803"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /clone="RP4-803K15"
                     /clone_lib="RPCI-4"
     gene            join(AL132765.38:146316. .146411,3171. .3333,5478. .5606,
                     16120. .16326)
                     /locus_tag="RP4-803K15.1-002"
     mRNA            join(AL132765.38:146316. .146411,3171. .3333,5478. .5606,
                     16120. .16326)
                     /locus_tag="RP4-803K15.1-002"
                     /product="chromosome 20 open reading frame 179"
                     /note="match: ESTs: Em:AI075928.1 Em:AI150321.1
                     Em:AI204602.1
                     match: cDNAs: Em:BC062735.1"
     gene            join(AL132765.38:146304. .146411,5478. .5606,16120. .16326)
                     /locus_tag="RP4-803K15.1-001"
     mRNA            join(AL132765.38:146304. .146411,5478. .5606,16120. .16326)
                     /locus_tag="RP4-803K15.1-001"
                     /product="chromosome 20 open reading frame 179"
                     /note="match: ESTs: Em:AA394107.1 Em:AA629341.1
                     Em:AA862555.1 Em:AI066604.1 Em:BG459855.1
                     match: cDNAs: Em:BC054871.1"
     gene            join(AL132765.38:146298. .146411,
                     AL132765.38:158517. .158896,5478. .5606)
                     /locus_tag="RP4-803K15.1-003"
     mRNA            join(AL132765.38:146298. .146411,
                     AL132765.38:158517. .158896,5478. .5606)
                     /locus_tag="RP4-803K15.1-003"
                     /product="chromosome 20 open reading frame 179"
                     /note="match: ESTs: Em:BU853192.1"
     CDS             join(5481. .5606,16120. .16266)
                     /locus_tag="RP4-803K15.1-002"
                     /standard_name="OTTHUMP00000040746"
                     /note="match: proteins: Sw:Q9H503 Tr:Q7Z4M6 Tr:Q8BVR0"
                     /codon_start=1
                     /product="chromosome 20 open reading frame 179"
                     /protein_id="CAC15540.1"
                     /db_xref="GI:11136903"
                     /db_xref="Genew:16172"
                     /db_xref="GOA:Q9H503"
                     /db_xref="HSSP:1CI4"
                     /db_xref="InterPro:IPR004122"
                     /db_xref="UniProt/Swiss-Prot:Q9H503"
                     /translation="MDNMSPRLRAFLSEPIGEKDVCWVDGISHELAINLVTKGINKAY
                     ILLGQFLLMHKNEAEFQRWLICCFGATECEAQQTSHCLKEWCACFL"
     gene            join(11634. .11963,16120. .16253)
                     /locus_tag="RP4-803K15.1-004"
     mRNA            join(11634. .11963,16120. .16253)
                     /locus_tag="RP4-803K15.1-004"
                     /product="chromosome 20 open reading frame 179"
                     /note="match: ESTs: Em:AA442269.1"
     polyA_site      16326
                     /locus_tag="RP4-803K15.1-001"
     misc_feature    51457
                     /note="Clone_left_end: RP5-905G11"
ORIGIN

Alignment Scores:

Pred. No.:                1.04e+04      Length:          51556
Score:                    42.00         Matches:         6
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       8             Gaps:            0

US-09-858-332G-28 (1-6) x AL160071 (1-51556)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      44832 CATAATCATAATCATAAT 44849

RESULT 447
AC108475
LOCUS       AC108475               51975 bp    DNA     linear   PRI 29-MAY-2002
DEFINITION  Homo sapiens BAC clone RP11-1365D11 from 4, complete sequence.
ACCESSION   AC108475
VERSION     AC108475.5  GI:20163175
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 51975)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
   PUBMED   9847074
REFERENCE   2  (bases 1 to 51975)
  AUTHORS   Desai,A., Kozlowicz,A., Creason,K. and Nguyen,C.
  TITLE     The sequence of Homo sapiens BAC clone RP11-1365D11
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 51975)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 51975)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 51975)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-APR-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   6  (bases 1 to 51975)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAY-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Apr 17, 2002 this sequence version replaced gi:20136994.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH1365D11
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
```

all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at http://www.chori.org
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is RP11-478C6, 2000 bp overlap. Actual start of this clone is at base position 1 of RP11-1365D11; actual end is at base position 141594 of RP11-478C6.

Single plasmid region exists between 24434 and 24458. Polymorphisms exist between AC079140 and AC108475.

FEATURES	Location/Qualifiers
source	1. .51975
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="4"
	/map="4"
	/clone="RP11-1365D11"
	/clone_lib="RPCI-11"
repeat_region	5228. .5377
	/rpt_family="L1"
repeat_region	5347. .5521
	/rpt_family="L1"
repeat_region	5587. .5769
	/rpt_family="MIR"
repeat_region	5917. .6215
	/rpt_family="Alu"
repeat_region	6612. .6908
	/rpt_family="ERV1"
repeat_region	6920. .7399
	/rpt_family="ERVK"
repeat_region	7732. .8301
	/rpt_family="ERV1"
repeat_region	8429. .8617
	/rpt_family="ERV1"
repeat_region	8995. .9143
	/rpt_family="Alu"
repeat_region	10326. .10788
	/rpt_family="ERV1"
repeat_region	10924. .11005
	/rpt_family="ERV1"
repeat_region	11006. .11220
	/rpt_family="ERV1"
repeat_region	11244. .11283
	/rpt_family="L1"
repeat_region	11284. .11580
	/rpt_family="Alu"
repeat_region	11581. .12646
	/rpt_family="L1"
repeat_region	13097. .13190
	/rpt_family="GA-rich"
repeat_region	13877. .14166
	/rpt_family="L1"
repeat_region	14167. .14447
	/rpt_family="Alu"
repeat_region	14448. .14783
	/rpt_family="L1"
repeat_region	15547. .15591
	/rpt_family="(TTTA)n"
repeat_region	15592. .15874
	/rpt_family="Alu"
repeat_region	20357. .20654
	/rpt_family="Alu"
repeat_region	21438. .21458
	/rpt_family="(CTGTG)n"
repeat_region	21995. .22201
	/rpt_family="L1"
repeat_region	22202. .22368
	/rpt_family="Alu"
repeat_region	22419. .23257
	/rpt_family="L1"
repeat_region	23291. .23591
	/rpt_family="Alu"
repeat_region	23903. .24185
	/rpt_family="Alu"
repeat_region	24287. .24593
	/rpt_family="Alu"
repeat_region	24905. .25686
	/rpt_family="ERV1"
repeat_region	25687. .25991
	/rpt_family="ERV1"
repeat_region	25968. .26005
	/rpt_family="ERV1"
repeat_region	26000. .26383
	/rpt_family="ERV1"
repeat_region	26384. .26407
	/rpt_family="(TTTCC)n"
repeat_region	26408. .26618
	/rpt_family="ERV1"
repeat_region	29391. .30085
	/rpt_family="ERV1"
repeat_region	30140. .30216
	/rpt_family="ERV1"
repeat_region	31010. .31077
	/rpt_family="(TGGA)n"
repeat_region	31974. .32169
	/rpt_family="MIR"
repeat_region	32237. .32352
	/rpt_family="MIR"
repeat_region	33381. .33708
	/rpt_family="Alu"
repeat_region	35136. .35498
	/rpt_family="MaLR"
repeat_region	36981. .37289
	/rpt_family="Alu"
repeat_region	37408. .37770
	/rpt_family="MaLR"
repeat_region	39205. .39382
	/rpt_family="(TG)n"
repeat_region	39413. .39581
	/rpt_family="(TG)n"
repeat_region	39594. .39656
	/rpt_family="(TG)n"
repeat_region	39982. .40042
	/rpt_family="MIR"
repeat_region	40192. .40307
	/rpt_family="MIR"
repeat_region	40945. .41908
	/rpt_family="ERVK"
repeat_region	42161. .42185
	/rpt_family="AT_rich"
repeat_region	42650. .42821

Alignment Scores:

Pred. No.:	1.05e+04	Length:	51975
Score:	42.00	Matches:	6

```
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     8              Gaps:          0

US-09-858-332G-28 (1-6) x AC108475 (1-51975)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      49420 CATAATCACAACCATAAT 49437

RESULT 448
AC017353/c
LOCUS       AC017353               54309 bp    DNA     linear   HTG 09-DEC-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC017353
VERSION     AC017353.1  GI:6553633
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 54309)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10210407 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .54309
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
ORIGIN

Alignment Scores:
Pred. No.:              1.1e+04        Length:        54309
Score:                  42.00          Matches:       6
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     14             Gaps:          0

US-09-858-332G-28 (1-6) x AC017353 (1-54309)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      32359 CACAATCATAACCATAAC 32342

RESULT 449
AC084717/c
LOCUS       AC084717               56310 bp    DNA     linear   HTG 15-JAN-2001
DEFINITION  Homo sapiens clone RP11-815P21, WORKING DRAFT SEQUENCE, 3 unordered
            pieces.
ACCESSION   AC084717
VERSION     AC084717.2  GI:12229387
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 56310)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-815P21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 56310)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
            FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
            Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
            Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G.,
            Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
            O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
            Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jan 15, 2001 this sequence version replaced gi:11120856.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L11499
                Center clone name: 815_P_21
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 55130 bases at least Q40
                Consensus quality: 55883 bases at least Q30
                Consensus quality: 56050 bases at least Q20
                Insert size: 61000; agarose-fp
                Insert size: 56110; sum-of-contigs
                Quality coverage: 7.9 in Q20 bases; agarose-fp
                Quality coverage: 8.6 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    46566: contig of 46566 bp in length
            *    46567    46666: gap of 100 bp
            *    46667    50756: contig of 4090 bp in length
            *    50757    50856: gap of 100 bp
            *    50857    56310: contig of 5454 bp in length.
FEATURES             Location/Qualifiers
     source          1. .56310
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-815P21"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .46566
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     gap             46567. .46666
                     /estimated_length=100
```

```
     misc_feature    46667. .50756
                     /note="assembly_fragment"
     gap             50757. .50856
                     /estimated_length=100
     misc_feature    50857. .56310
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
ORIGIN

Alignment Scores:
Pred. No.:                1.13e+04       Length:         56310
Score:                    42.00          Matches:        6
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       14             Gaps:           0

US-09-858-332G-28 (1-6) x AC084717 (1-56310)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      14375 CATAATCATAATCATAAT 14358

RESULT 450
AC019800/c
LOCUS       AC019800               56366 bp    DNA     linear   HTG 03-JAN-2000
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC019800
VERSION     AC019800.1  GI:6665097
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 56366)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10210991 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .56366
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
ORIGIN

Alignment Scores:
Pred. No.:                1.14e+04       Length:         56366
Score:                    42.00          Matches:        6
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       14             Gaps:           0

US-09-858-332G-28 (1-6) x AC019800 (1-56366)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      37086 CACAATCATAATCACAAT 37069

RESULT 451
AE003846_3
WPCOMMENT
  Sequence split into 4 fragments  LOCUS AE003846 Accession AE003846
    Fragment Name      Begin     End
     AE003846_0            1    110000
     AE003846_1       100001    210000
     AE003846_2       200001    310000
     AE003846_3       300001    356494
Continuation (4 of 4) of AE003846 from base 300001 (AE003846 Drosophila melanogaster chro

Alignment Scores:
Pred. No.:                1.14e+04       Length:         56494
Score:                    42.00          Matches:        6
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       2              Gaps:           0

US-09-858-332G-28 (1-6) x AE003846_3 (1-56494)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      14277 CATAATCATAATCATAAT 14294

RESULT 452
AC145220/c
LOCUS       AC145220               57331 bp    DNA     linear   HTG 04-AUG-2005
DEFINITION  Medicago truncatula clone mth2-24j17, WORKING DRAFT SEQUENCE, 5
            unordered pieces.
ACCESSION   AC145220
VERSION     AC145220.12  GI:71795750
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 57331)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Medicago truncatula BAC Clone mth2-24j17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 57331)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUN-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 57331)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-AUG-2005) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Aug 4, 2005 this sequence version replaced gi:67010125.
            -------------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ----------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2062: contig of 2062 bp in length
            *     2063     2162: gap of unknown length
            *     2163     4251: contig of 2089 bp in length
            *     4252     4351: gap of unknown length
            *     4352    25744: contig of 21393 bp in length
            *    25745    25844: gap of unknown length
```

```
            *    25845    33075: contig of 7231 bp in length
            *    33076    33175: gap of unknown length
            *    33176    57331: contig of 24156 bp in length.
FEATURES             Location/Qualifiers
     source          1. .57331
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3880"
                     /clone="mth2-24j17"
                     /clone_lib="Medicago truncatula BAC library H2"
     gap             2063. .2162
                     /estimated_length=unknown
     gap             4252. .4351
                     /estimated_length=unknown
     gap             25745. .25844
                     /estimated_length=unknown
     gap             33076. .33175
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              1.15e+04        Length:         57331
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AC145220 (1-57331)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       8034 CATAATCATAATCATAAT 8017

RESULT 453
CR759870
LOCUS       CR759870               57743 bp    DNA     linear   PRI 14-NOV-2004
DEFINITION  Human DNA sequence from clone DADB-225M10 on chromosome 6, complete
            sequence.
ACCESSION   CR759870
VERSION     CR759870.6  GI:55734208
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 57743)
  AUTHORS   Brown,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-NOV-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 14, 2004 this sequence version replaced gi:55468522.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            DADB-225M10 is from a DNA-arts DBB human bac library VECTOR:
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .57743
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="DADB-225M10"
                     /clone_lib="DNA-arts-BAC.1-DBB.1"
ORIGIN

Alignment Scores:
Pred. No.:              1.16e+04        Length:         57743
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     8               Gaps:           0

US-09-858-332G-28 (1-6) x CR759870 (1-57743)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      14113 CATAATCATAATCATAAT 14130

RESULT 454
AC103926
LOCUS       AC103926               59605 bp    DNA     linear   HTG 01-DEC-2001
DEFINITION  Mus musculus clone RP23-54C6, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC103926
VERSION     AC103926.1  GI:17223212
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 59605)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-54C6
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 59605)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
```

```
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-DEC-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L14228
                Center clone name: 54_C_6
            --------------.
            * NOTE: This record contains 78 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      657: contig of 657 bp in length
            *      658      757: gap of 100 bp
            *      758     1437: contig of 680 bp in length
            *     1438     1537: gap of 100 bp
            *     1538     2225: contig of 688 bp in length
            *     2226     2325: gap of 100 bp
            *     2326     3014: contig of 689 bp in length
            *     3015     3114: gap of 100 bp
            *     3115     3769: contig of 655 bp in length
            *     3770     3869: gap of 100 bp
            *     3870     4498: contig of 629 bp in length
            *     4499     4598: gap of 100 bp
            *     4599     5268: contig of 670 bp in length
            *     5269     5368: gap of 100 bp
            *     5369     6048: contig of 680 bp in length
            *     6049     6148: gap of 100 bp
            *     6149     6807: contig of 659 bp in length
            *     6808     6907: gap of 100 bp
            *     6908     7560: contig of 653 bp in length
            *     7561     7660: gap of 100 bp
            *     7661     8342: contig of 682 bp in length
            *     8343     8442: gap of 100 bp
            *     8443     9084: contig of 642 bp in length
            *     9085     9184: gap of 100 bp
            *     9185     9876: contig of 692 bp in length
            *     9877     9976: gap of 100 bp
            *     9977    10648: contig of 672 bp in length
            *    10649    10748: gap of 100 bp
            *    10749    11433: contig of 685 bp in length
            *    11434    11533: gap of 100 bp
            *    11534    12164: contig of 631 bp in length
            *    12165    12264: gap of 100 bp
            *    12265    12930: contig of 666 bp in length
            *    12931    13030: gap of 100 bp
            *    13031    13679: contig of 649 bp in length
            *    13680    13779: gap of 100 bp
            *    13780    14437: contig of 658 bp in length
            *    14438    14537: gap of 100 bp
            *    14538    15203: contig of 666 bp in length
            *    15204    15303: gap of 100 bp
            *    15304    16006: contig of 703 bp in length
            *    16007    16106: gap of 100 bp
            *    16107    16806: contig of 700 bp in length
            *    16807    16906: gap of 100 bp
            *    16907    17559: contig of 653 bp in length
            *    17560    17659: gap of 100 bp
            *    17660    18333: contig of 674 bp in length
            *    18334    18433: gap of 100 bp
            *    18434    19107: contig of 674 bp in length
            *    19108    19207: gap of 100 bp
            *    19208    19893: contig of 686 bp in length
            *    19894    19993: gap of 100 bp
            *    19994    20656: contig of 663 bp in length
            *    20657    20756: gap of 100 bp
            *    20757    21419: contig of 663 bp in length
            *    21420    21519: gap of 100 bp
            *    21520    22174: contig of 655 bp in length
            *    22175    22274: gap of 100 bp
            *    22275    22939: contig of 665 bp in length
            *    22940    23039: gap of 100 bp
            *    23040    23703: contig of 664 bp in length
            *    23704    23803: gap of 100 bp
            *    23804    24472: contig of 669 bp in length
            *    24473    24572: gap of 100 bp
            *    24573    25234: contig of 662 bp in length
            *    25235    25334: gap of 100 bp
            *    25335    26013: contig of 679 bp in length
            *    26014    26113: gap of 100 bp
            *    26114    26761: contig of 648 bp in length
            *    26762    26861: gap of 100 bp
            *    26862    27528: contig of 667 bp in length
            *    27529    27628: gap of 100 bp
            *    27629    28280: contig of 652 bp in length
            *    28281    28380: gap of 100 bp
            *    28381    29028: contig of 648 bp in length
            *    29029    29128: gap of 100 bp
            *    29129    29787: contig of 659 bp in length
            *    29788    29887: gap of 100 bp
            *    29888    30540: contig of 653 bp in length
            *    30541    30640: gap of 100 bp
            *    30641    31321: contig of 681 bp in length
            *    31322    31421: gap of 100 bp
            *    31422    32105: contig of 684 bp in length
            *    32106    32205: gap of 100 bp
            *    32206    32863: contig of 658 bp in length
            *    32864    32963: gap of 100 bp
            *    32964    33650: contig of 687 bp in length
            *    33651    33750: gap of 100 bp
            *    33751    34428: contig of 678 bp in length
            *    34429    34528: gap of 100 bp
            *    34529    35222: contig of 694 bp in length
            *    35223    35322: gap of 100 bp
            *    35323    35983: contig of 661 bp in length
            *    35984    36083: gap of 100 bp
            *    36084    36740: contig of 657 bp in length
            *    36741    36840: gap of 100 bp
            *    36841    37498: contig of 658 bp in length
            *    37499    37598: gap of 100 bp
            *    37599    38266: contig of 668 bp in length
            *    38267    38366: gap of 100 bp
            *    38367    38996: contig of 630 bp in length
            *    38997    39096: gap of 100 bp
            *    39097    39743: contig of 647 bp in length
            *    39744    39843: gap of 100 bp
            *    39844    40519: contig of 676 bp in length
            *    40520    40619: gap of 100 bp
            *    40620    41278: contig of 659 bp in length
            *    41279    41378: gap of 100 bp
            *    41379    42055: contig of 677 bp in length
            *    42056    42155: gap of 100 bp
            *    42156    42784: contig of 629 bp in length
            *    42785    42884: gap of 100 bp
            *    42885    43538: contig of 654 bp in length
            *    43539    43638: gap of 100 bp
            *    43639    44294: contig of 656 bp in length
            *    44295    44394: gap of 100 bp
            *    44395    45061: contig of 667 bp in length
            *    45062    45161: gap of 100 bp
```

```
            *    45162    45837: contig of 676 bp in length
            *    45838    45937: gap of 100 bp
            *    45938    46616: contig of 679 bp in length
            *    46617    46716: gap of 100 bp
            *    46717    47379: contig of 663 bp in length
            *    47380    47479: gap of 100 bp
            *    47480    48160: contig of 681 bp in length
            *    48161    48260: gap of 100 bp
            *    48261    48933: contig of 673 bp in length
            *    48934    49033: gap of 100 bp
            *    49034    49723: contig of 690 bp in length
            *    49724    49823: gap of 100 bp
            *    49824    50453: contig of 630 bp in length
            *    50454    50553: gap of 100 bp
            *    50554    51200: contig of 647 bp in length
            *    51201    51300: gap of 100 bp
            *    51301    51957: contig of 657 bp in length
            *    51958    52057: gap of 100 bp
            *    52058    52718: contig of 661 bp in length

Alignment Scores:
Pred. No.:                1.2e+04        Length:          59605
Score:                    42.00          Matches:         6
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       14             Gaps:            0

US-09-858-332G-28 (1-6) x AC103926 (1-59605)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      17018 CACAACCACAACCACAAC 17035

RESULT 455
AC103926/c
LOCUS       AC103926               59605 bp    DNA     linear   HTG 01-DEC-2001
DEFINITION  Mus musculus clone RP23-54C6, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC103926
VERSION     AC103926.1  GI:17223212
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 59605)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-54C6
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 59605)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-DEC-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L14228
                Center clone name: 54_C_6
            --------------.
            * NOTE: This record contains 78 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      657: contig of 657 bp in length
            *      658      757: gap of 100 bp
            *      758     1437: contig of 680 bp in length
            *     1438     1537: gap of 100 bp
            *     1538     2225: contig of 688 bp in length
            *     2226     2325: gap of 100 bp
            *     2326     3014: contig of 689 bp in length
            *     3015     3114: gap of 100 bp
            *     3115     3769: contig of 655 bp in length
            *     3770     3869: gap of 100 bp
            *     3870     4498: contig of 629 bp in length
            *     4499     4598: gap of 100 bp
            *     4599     5268: contig of 670 bp in length
            *     5269     5368: gap of 100 bp
            *     5369     6048: contig of 680 bp in length
            *     6049     6148: gap of 100 bp
            *     6149     6807: contig of 659 bp in length
            *     6808     6907: gap of 100 bp
            *     6908     7560: contig of 653 bp in length
            *     7561     7660: gap of 100 bp
            *     7661     8342: contig of 682 bp in length
            *     8343     8442: gap of 100 bp
            *     8443     9084: contig of 642 bp in length
            *     9085     9184: gap of 100 bp
            *     9185     9876: contig of 692 bp in length
            *     9877     9976: gap of 100 bp
            *     9977    10648: contig of 672 bp in length
            *    10649    10748: gap of 100 bp
            *    10749    11433: contig of 685 bp in length
            *    11434    11533: gap of 100 bp
            *    11534    12164: contig of 631 bp in length
            *    12165    12264: gap of 100 bp
            *    12265    12930: contig of 666 bp in length
            *    12931    13030: gap of 100 bp
            *    13031    13679: contig of 649 bp in length
            *    13680    13779: gap of 100 bp
            *    13780    14437: contig of 658 bp in length
            *    14438    14537: gap of 100 bp
            *    14538    15203: contig of 666 bp in length
            *    15204    15303: gap of 100 bp
            *    15304    16006: contig of 703 bp in length
            *    16007    16106: gap of 100 bp
            *    16107    16806: contig of 700 bp in length
            *    16807    16906: gap of 100 bp
            *    16907    17559: contig of 653 bp in length
            *    17560    17659: gap of 100 bp
            *    17660    18333: contig of 674 bp in length
```

```
*    18334    18433: gap of 100 bp
*    18434    19107: contig of 674 bp in length
*    19108    19207: gap of 100 bp
*    19208    19893: contig of 686 bp in length
*    19894    19993: gap of 100 bp
*    19994    20656: contig of 663 bp in length
*    20657    20756: gap of 100 bp
*    20757    21419: contig of 663 bp in length
*    21420    21519: gap of 100 bp
*    21520    22174: contig of 655 bp in length
*    22175    22274: gap of 100 bp
*    22275    22939: contig of 665 bp in length
*    22940    23039: gap of 100 bp
*    23040    23703: contig of 664 bp in length
*    23704    23803: gap of 100 bp
*    23804    24472: contig of 669 bp in length
*    24473    24572: gap of 100 bp
*    24573    25234: contig of 662 bp in length
*    25235    25334: gap of 100 bp
*    25335    26013: contig of 679 bp in length
*    26014    26113: gap of 100 bp
*    26114    26761: contig of 648 bp in length
*    26762    26861: gap of 100 bp
*    26862    27528: contig of 667 bp in length
*    27529    27628: gap of 100 bp
*    27629    28280: contig of 652 bp in length
*    28281    28380: gap of 100 bp
*    28381    29028: contig of 648 bp in length
*    29029    29128: gap of 100 bp
*    29129    29787: contig of 659 bp in length
*    29788    29887: gap of 100 bp
*    29888    30540: contig of 653 bp in length
*    30541    30640: gap of 100 bp
*    30641    31321: contig of 681 bp in length
*    31322    31421: gap of 100 bp
*    31422    32105: contig of 684 bp in length
*    32106    32205: gap of 100 bp
*    32206    32863: contig of 658 bp in length
*    32864    32963: gap of 100 bp
*    32964    33650: contig of 687 bp in length
*    33651    33750: gap of 100 bp
*    33751    34428: contig of 678 bp in length
*    34429    34528: gap of 100 bp
*    34529    35222: contig of 694 bp in length
*    35223    35322: gap of 100 bp
*    35323    35983: contig of 661 bp in length
*    35984    36083: gap of 100 bp
*    36084    36740: contig of 657 bp in length
*    36741    36840: gap of 100 bp
*    36841    37498: contig of 658 bp in length
*    37499    37598: gap of 100 bp
*    37599    38266: contig of 668 bp in length
*    38267    38366: gap of 100 bp
*    38367    38996: contig of 630 bp in length
*    38997    39096: gap of 100 bp
*    39097    39743: contig of 647 bp in length
*    39744    39843: gap of 100 bp
*    39844    40519: contig of 676 bp in length
*    40520    40619: gap of 100 bp
*    40620    41278: contig of 659 bp in length
*    41279    41378: gap of 100 bp
*    41379    42055: contig of 677 bp in length
*    42056    42155: gap of 100 bp
*    42156    42784: contig of 629 bp in length
*    42785    42884: gap of 100 bp
*    42885    43538: contig of 654 bp in length
*    43539    43638: gap of 100 bp
*    43639    44294: contig of 656 bp in length
*    44295    44394: gap of 100 bp
*    44395    45061: contig of 667 bp in length
*    45062    45161: gap of 100 bp
*    45162    45837: contig of 676 bp in length
*    45838    45937: gap of 100 bp
*    45938    46616: contig of 679 bp in length
*    46617    46716: gap of 100 bp
*    46717    47379: contig of 663 bp in length
*    47380    47479: gap of 100 bp
*    47480    48160: contig of 681 bp in length
*    48161    48260: gap of 100 bp
*    48261    48933: contig of 673 bp in length
*    48934    49033: gap of 100 bp
*    49034    49723: contig of 690 bp in length
*    49724    49823: gap of 100 bp
*    49824    50453: contig of 630 bp in length
*    50454    50553: gap of 100 bp
*    50554    51200: contig of 647 bp in length
*    51201    51300: gap of 100 bp
*    51301    51957: contig of 657 bp in length
*    51958    52057: gap of 100 bp
*    52058    52718: contig of 661 bp in length

Alignment Scores:
Pred. No.:                1.2e+04        Length:          59605
Score:                    42.00          Matches:         6
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       14             Gaps:            0

US-09-858-332G-28 (1-6) x AC103926 (1-59605)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      52410 CACAACCACAACCACAAC 52393

RESULT 456
AC149161
LOCUS       AC149161               60038 bp    DNA     linear   HTG 20-MAY-2004
DEFINITION  Xenopus tropicalis clone ISB-33F19, WORKING DRAFT SEQUENCE, 5
            unordered pieces.
ACCESSION   AC149161
VERSION     AC149161.2  GI:47524008
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Xenopus tropicalis (Silurana tropicalis)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 60038)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Unpublished
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 60038)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAY-2004) Production Genomics Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA
            94598-1698, USA
REFERENCE   3  (bases 1 to 60038)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAY-2004) Production Genomics Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA
            94598-1698, USA
COMMENT     On May 20, 2004 this sequence version replaced gi:47498114.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 2785762
            Center clone name: ISB-33F19
            -------
            Summary Statistics
            Consensus quality: 56990 bases at least Q40
```

```
            Consensus quality: 58567 bases at least Q30
            Consensus quality: 59294 bases at least Q20
            Estimated insert size: 70000; agarose-fp estimation
            Estimated insert size: 59638; sum-of-contigs estimation
            Quality coverage:  8.43 in Q20 bases; agarose-fp estimation
            Quality coverage:  9.9 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     3956: contig of 3956 bp in length
            *    3957     4056: gap of unknown length
            *    4057    10168: contig of 6112 bp in length
            *   10169    10268: gap of unknown length
            *   10269    23159: contig of 12891 bp in length
            *   23160    23259: gap of unknown length
            *   23260    35940: contig of 12681 bp in length
            *   35941    36040: gap of unknown length
            *   36041    60038: contig of 23998 bp in length.
FEATURES             Location/Qualifiers
     source          1. .60038
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:8364"
                     /clone="ISB-33F19"
     gap             3957. .4056
                     /estimated_length=unknown
     gap             10169. .10268
                     /estimated_length=unknown
     gap             23160. .23259
                     /estimated_length=unknown
     gap             35941. .36040
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              1.21e+04        Length:         60038
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AC149161 (1-60038)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      42128 CATAACCATAACCATAAT 42145

RESULT 457
AC149161/c
LOCUS       AC149161               60038 bp    DNA     linear   HTG 20-MAY-2004
DEFINITION  Xenopus tropicalis clone ISB-33F19, WORKING DRAFT SEQUENCE, 5
            unordered pieces.
ACCESSION   AC149161
VERSION     AC149161.2  GI:47524008
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Xenopus tropicalis (Silurana tropicalis)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 60038)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Unpublished
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 60038)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAY-2004) Production Genomics Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA
            94598-1698, USA
REFERENCE   3  (bases 1 to 60038)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAY-2004) Production Genomics Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA
            94598-1698, USA
COMMENT     On May 20, 2004 this sequence version replaced gi:47498114.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 2785762
            Center clone name: ISB-33F19
            -------
            Summary Statistics
            Consensus quality: 56990 bases at least Q40
            Consensus quality: 58567 bases at least Q30
            Consensus quality: 59294 bases at least Q20
            Estimated insert size: 70000; agarose-fp estimation
            Estimated insert size: 59638; sum-of-contigs estimation
            Quality coverage:  8.43 in Q20 bases; agarose-fp estimation
            Quality coverage:  9.9 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     3956: contig of 3956 bp in length
            *    3957     4056: gap of unknown length
            *    4057    10168: contig of 6112 bp in length
            *   10169    10268: gap of unknown length
            *   10269    23159: contig of 12891 bp in length
            *   23160    23259: gap of unknown length
            *   23260    35940: contig of 12681 bp in length
            *   35941    36040: gap of unknown length
            *   36041    60038: contig of 23998 bp in length.
FEATURES             Location/Qualifiers
     source          1. .60038
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:8364"
                     /clone="ISB-33F19"
     gap             3957. .4056
                     /estimated_length=unknown
     gap             10169. .10268
                     /estimated_length=unknown
     gap             23160. .23259
                     /estimated_length=unknown
     gap             35941. .36040
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              1.21e+04        Length:         60038
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AC149161 (1-60038)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      34420 CATAACCATAACCATAAT 34403
```

```
RESULT 458
AC073922/c
LOCUS       AC073922               60713 bp    DNA     linear   HTG 15-NOV-2000
DEFINITION  Homo sapiens chromosome 15 clone RP11-173P23 map 15, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC073922
VERSION     AC073922.3  GI:11178123
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 60713)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-173P23
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 60713)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUL-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 15, 2000 this sequence version replaced gi:11120864.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L1279
                Center clone name: 173_P_23
            --------------.
            * NOTE: This record contains 76 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      703: contig of 703 bp in length
            *      704      803: gap of 100 bp
            *      804     1497: contig of 694 bp in length
            *     1498     1597: gap of 100 bp
            *     1598     2306: contig of 709 bp in length
            *     2307     2406: gap of 100 bp
            *     2407     3090: contig of 684 bp in length
            *     3091     3190: gap of 100 bp
            *     3191     3867: contig of 677 bp in length
            *     3868     3967: gap of 100 bp
            *     3968     4646: contig of 679 bp in length
            *     4647     4746: gap of 100 bp
            *     4747     5445: contig of 699 bp in length
            *     5446     5545: gap of 100 bp
            *     5546     6251: contig of 706 bp in length
            *     6252     6351: gap of 100 bp
            *     6352     7062: contig of 711 bp in length
            *     7063     7162: gap of 100 bp
            *     7163     7874: contig of 712 bp in length
            *     7875     7974: gap of 100 bp
            *     7975     8690: contig of 716 bp in length
            *     8691     8790: gap of 100 bp
            *     8791     9479: contig of 689 bp in length
            *     9480     9579: gap of 100 bp
            *     9580    10280: contig of 701 bp in length
            *    10281    10380: gap of 100 bp
            *    10381    11087: contig of 707 bp in length
            *    11088    11187: gap of 100 bp
            *    11188    11890: contig of 703 bp in length
            *    11891    11990: gap of 100 bp
            *    11991    12674: contig of 684 bp in length
            *    12675    12774: gap of 100 bp
            *    12775    13488: contig of 714 bp in length
            *    13489    13588: gap of 100 bp
            *    13589    14289: contig of 701 bp in length
            *    14290    14389: gap of 100 bp
            *    14390    15090: contig of 701 bp in length
            *    15091    15190: gap of 100 bp
            *    15191    15896: contig of 706 bp in length
            *    15897    15996: gap of 100 bp
            *    15997    16721: contig of 725 bp in length
            *    16722    16821: gap of 100 bp
            *    16822    17499: contig of 678 bp in length
            *    17500    17599: gap of 100 bp
            *    17600    18285: contig of 686 bp in length
            *    18286    18385: gap of 100 bp
            *    18386    19090: contig of 705 bp in length
            *    19091    19190: gap of 100 bp
            *    19191    19895: contig of 705 bp in length
            *    19896    19995: gap of 100 bp
            *    19996    20686: contig of 691 bp in length
            *    20687    20786: gap of 100 bp
            *    20787    21500: contig of 714 bp in length
            *    21501    21600: gap of 100 bp
            *    21601    22310: contig of 710 bp in length
            *    22311    22410: gap of 100 bp
            *    22411    23133: contig of 723 bp in length
            *    23134    23233: gap of 100 bp
            *    23234    23936: contig of 703 bp in length
            *    23937    24036: gap of 100 bp
            *    24037    24724: contig of 688 bp in length
            *    24725    24824: gap of 100 bp
            *    24825    25518: contig of 694 bp in length
            *    25519    25618: gap of 100 bp
            *    25619    26350: contig of 732 bp in length
            *    26351    26450: gap of 100 bp
            *    26451    27154: contig of 704 bp in length
            *    27155    27254: gap of 100 bp
            *    27255    27955: contig of 701 bp in length
            *    27956    28055: gap of 100 bp
            *    28056    28762: contig of 707 bp in length
            *    28763    28862: gap of 100 bp
            *    28863    29555: contig of 693 bp in length
            *    29556    29655: gap of 100 bp
            *    29656    30347: contig of 692 bp in length
            *    30348    30447: gap of 100 bp
            *    30448    31133: contig of 686 bp in length
            *    31134    31233: gap of 100 bp
            *    31234    31935: contig of 702 bp in length
```

```
          *    31936    32035: gap of 100 bp
          *    32036    32724: contig of 689 bp in length
          *    32725    32824: gap of 100 bp
          *    32825    33509: contig of 685 bp in length
          *    33510    33609: gap of 100 bp
          *    33610    34284: contig of 675 bp in length
          *    34285    34384: gap of 100 bp
          *    34385    35091: contig of 707 bp in length
          *    35092    35191: gap of 100 bp
          *    35192    35897: contig of 706 bp in length
          *    35898    35997: gap of 100 bp
          *    35998    36709: contig of 712 bp in length
          *    36710    36809: gap of 100 bp
          *    36810    37520: contig of 711 bp in length
          *    37521    37620: gap of 100 bp
          *    37621    38298: contig of 678 bp in length
          *    38299    38398: gap of 100 bp
          *    38399    39120: contig of 722 bp in length
          *    39121    39220: gap of 100 bp
          *    39221    39922: contig of 702 bp in length
          *    39923    40022: gap of 100 bp
          *    40023    40719: contig of 697 bp in length
          *    40720    40819: gap of 100 bp
          *    40820    41467: contig of 648 bp in length
          *    41468    41567: gap of 100 bp
          *    41568    42286: contig of 719 bp in length
          *    42287    42386: gap of 100 bp
          *    42387    43082: contig of 696 bp in length
          *    43083    43182: gap of 100 bp
          *    43183    43877: contig of 695 bp in length
          *    43878    43977: gap of 100 bp
          *    43978    44695: contig of 718 bp in length
          *    44696    44795: gap of 100 bp
          *    44796    45490: contig of 695 bp in length
          *    45491    45590: gap of 100 bp
          *    45591    46271: contig of 681 bp in length
          *    46272    46371: gap of 100 bp
          *    46372    47076: contig of 705 bp in length
          *    47077    47176: gap of 100 bp
          *    47177    47874: contig of 698 bp in length
          *    47875    47974: gap of 100 bp
          *    47975    48670: contig of 696 bp in length
          *    48671    48770: gap of 100 bp
          *    48771    49482: contig of 712 bp in length
          *    49483    49582: gap of 100 bp
          *    49583    50282: contig of 700 bp in length
          *    50283    50382: gap of 100 bp
          *    50383    51085: contig of 703 bp in length
          *    51086    51185: gap of 100 bp
          *    51186    51873: contig of 688 bp in length
          *    51874    51973: gap of 100 bp
          *    51974    52677: contig of 704 bp in length
          *    52678    52777: gap of 100 bp
          *    52778    53470: contig of 693 bp in length
          *    53471    53570: gap of 100 bp
          *    53571    54273: contig of 703 bp in length
          *    54274    54373: gap of 100 bp

Alignment Scores:
Pred. No.:              1.22e+04       Length:         60713
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     14             Gaps:           0

US-09-858-332G-28 (1-6) x AC073922 (1-60713)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      13846 CATAATCATAATCATAAT 13829

RESULT 459
AC017150/c
LOCUS       AC017150               60919 bp    DNA     linear   HTG 09-DEC-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC017150
VERSION     AC017150.1  GI:6553836
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 60919)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10209656 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .60919
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
ORIGIN

Alignment Scores:
Pred. No.:              1.22e+04       Length:         60919
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     14             Gaps:           0

US-09-858-332G-28 (1-6) x AC017150 (1-60919)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      37379 CATAATCATAATCATAAT 37362

RESULT 460
AC123513
LOCUS       AC123513               61052 bp    DNA     linear   HTG 29-MAY-2002
DEFINITION  Dictyostelium discoideum strain AX4 chromosome 2 map
            2779865-2840915, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC123513
VERSION     AC123513.1  GI:21240650
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 61052)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and Analysis of Chromosome 2 of Dictyostelium
  JOURNAL   Unpublished
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 61052)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAY-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     CDS predictions from GeneID may contain errors. Further Information
            is available from IMB Jena, Department of Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
```

```
            Agency : Deutsche Forschungsgemeinschaft (DFG).
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES              Location/Qualifiers
     source           1. .61052
                      /organism="Dictyostelium discoideum"
                      /mol_type="genomic DNA"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /chromosome="2"
                      /map="2779865-2840915"
     CDS              join(448. .628,705. .803,875. .1134)
                      /note="ORF_ID:dd_02708"
                      /pseudo
                      /codon_start=1
     CDS              complement(join(1193. .1583,1653. .6117,6198. .6330))
                      /note="ORF_ID:dd_02705"
                      /codon_start=1
                      /product="ABC1 TRANSPORTER"
                      /protein_id="AAM44362.1"
                      /db_xref="GI:21240651"
                      /translation="MGEKSSQLKTLLKKNLLLKSKSKCGICCEIVFPIIIVLLVFAIL
                      VLVQLFKPNYDLIKTTQFSNRINENNIIIYGGKAGSLNVEQKGVIDMMKFQLSNELNK
                      SISEVDVYFKEINDKSEMENYFQINSTQVFGGIWFESNQLSSVANTTNTNTTTNSFKY
                      SIRLDSNSVLDSNKVKDNGIDSSDYLTKNWAYIQIAMDQAIFGYFGLDYKLVINGQRY
                      PDPYVELWQKWINGRESVFKSAGSVFVSAALLIFTFRLVTELVVEKETKIREGMSIMG
                      LNQYCYFISWIITSLVTALPIDLIIIVILKGSQVIHSTSWIIVIVTLILYLLTLQLLA
                      FIFSMFFDKSKFAGLLTFLTILLINICGIFIGEYEIININIHIKLLLCCIFSPIGIACSF
                      YIMSIRDIPDTISTINLNQVISEKQIIGTFVFNIIFYTFLIWYLDKIVKTEYGTKEPW
                      YFLFTKRYWTVGAGKINKKNKNKEYNYNDIESTIQNNNNDNIEMVPIEVRNKTTISIR
                      NLRKEFKTGDGLRIAVNDLNLDMFDGQIHGLLGPNGSGKSTTISMLTGLLSPTSGTAF
                      ILGNDITYQMNEIRKCTGVCLQTDIIWNQLTVLEHLEIYASLKGITNKKQIKSESLKM
                      ANEIDLGEKLHTPAGSLSGGQKRKLCLGIAFIGRSTIIFLDEVSSGMDPSSRRKVWDF
                      LLSHKKGKTIILTTHYLDEADYLSDRISIISHGKLITDGSSLFLKNKYGVGYLLTCSK
                      SLNTIDQFNVDQVTQFIRDQIPDVTVLSNAGSEISFRLPTASLPVFSKFFKDFDENLS
                      NFHIDSYGISVTTLEEVFLKIGTDIDTDALSIADGNGNDDELKRSIGVSSTGIKSKQQ
                      LKALLIKRVKTSSKDFKAFLLSLLLPLLVIIGSIIVFKEVDNEVIFYNNSTEPLTFSL
                      LQQYGKNDIVPIQLAGTTSEFDFNKFLSNSPYFNQFQYLNNSINFNQYLINNYKQSSG
                      SINFTIPLSSSLINNTTKVISYNSLFNFNYIHSWPVHVNLINDALLRNHNGIGIECTS
                      LPFDHILTSFQKASQGMNIQSIVYFIVIMMAGFSLMAGSFAGSISQERTNRIKRLLYI
                      SGCKKYIYWVSNLLWDYFFAFILLLITCIILAIVDDKFNDQFGLFFLSLVLFSLSIIP
                      LSYLFSFKFSSFGKSTGAITAIHFSIGVIMTIAMIILRIEVIIKNSSSLENIADIIDI
                      VFNILSPLYAFSRVIFIISGFPGSLRLGALKIDDYWSLHYASTPIIILSVHVVVWTIF
                      ILLIDYSPEFKGYLRNPKTILPPPPPIDEDSDVSAERIRLESMSPIPSTDQGSGGGGD
                      MLQYKGLHKLFIGKGKNPNKNAVYNSTLGIPQGQTFGLLGLNGAGKTTTVSMLAGDIL
                      PTSGQITINGHDLVTDRAQALRGVSVCPQFDALITLLTAREQLSLYCAIKGVPDDKIS
                      LVVEAFIKMMDLGKIANSNTGGYSGGNKRKVSLSIAMLGNPSVVLIDEASSGCDPIIR
                      YRQCQVISELGKNKVIILTSHSMSEIQALCSRMTIMRDGQFKCLGSTQHIKSKFGAGY
                      SVEVKFKKSCLEVGIPQSLQCVLECFPNATILDQHDLMASLELPNPPENPIKVSEIFN
                      ILSTELSSILDDYSVSQTSLEQVFLKLTGATHEDRLNLNNQQQQQQTIPNSD"
     CDS              complement(join(7575. .8146,8235. .8364))
                      /note="ORF_ID:dd_02702"
                      /codon_start=1
                      /product="hypothetical protein"
                      /protein_id="AAM44363.1"
                      /db_xref="GI:21240652"
                      /translation="MEIKYFLVLLVGFLLVLPSIVNPYRKGVTITNQQPKINIYQLDV
                      NNPHNPNNNPHNPHNPNNNPHHPHHLHHHHHHHHHHHHHHHHHHHHHHHHHHPHHPHH
                      HPHHHHHPHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHPHHHPH
                      PHPHPHPHPHLHPNPHPHPHPHPHPHPHPHHHPNPNPHPHPHPHPHHHHHQEASECL
                      SYQGNRNRLFIKRDY"
     CDS              9039. .10208
                      /note="ORF_ID:dd_02724"
                      /codon_start=1
                      /product="POTASSIUM CHANNEL REGULATORY FACTOR"
                      /protein_id="AAM44364.1"
                      /db_xref="GI:21240653"
                      /translation="MSSLSNLKEQLEKKRKEAQLKNSNNNNNNNNNSNSNNGCDKSE
                      EPATKKYKTRGEIDKELKQKKEEERIKKEKEKEKEKEKENYNNNNNSNVNSNKNEN
                      NINIDNDNDSNDKINKLKENEGFLPDIVVIQRLRERSQPITYFGETSIDRAMRLKKLE
                      ENDPIEYTQGDNEFANILKQIEQRKMKKHQDSADNEKEKEENENNDKKKQQQQQENEE
                      SEEQELERLRIDSSKCKESKILYFLKGLLNDWITNLENRSDDEKKSRQGKIAEATYYQ
                      CKSHIQPLFTSLREKTLSIDILEHINEIVEFCKSREYVRANDEYLQMAIGNAPWPMGV
                      TMVGIHERSSREKIFSNQVAHVLNDEMQRKYIQAVKRLMTFCQTQYPTAPSRSIG"
     CDS              join(11612. .11806,12045. .12147,12280. .12424,12526. .12693,
                      12802. .13164,13244. .13607)
                      /note="ORF_ID:dd_02721"
                      /codon_start=1
                      /product="Putative Chromosome segregation ATPase"
                      /protein_id="AAM44365.1"
                      /db_xref="GI:21240654"
                      /translation="MNFRELDTSNGVKAAESSFRTKVVLAREIPWNGFASSNSITSEQ
                      YNLISKYDKHTDAEKKEKFAANSASYVNFFVNFINSTSNIEIIQYLLTLINEIIEIDP
                      RAAGAPSKITKDDDKSYPYSVFFRLLNREDAYTNLHASIALAQIMCAGKPTQNDVESF
                      FNWILKLLRKNNSSEVEVGLIALQSLLLKDDFRIFFNNIDGSALLLNILQALSTSSVN
                      IQLLYETIYAIWLLTYNKDIAAAYSGTGLVANLVQLVKTVAKEKIVRLSLSTLRNLLN
                      NGKSNEEMIDNGFVRMLNILNIKKWGDDDIPADIEVLINGLAKDIDNMSSFNKYKTEI
                      ISGELEWTPVHKSERFWKENISKFEENNYQVIKHLHQILKTSQSTPLQLSIACHDLCE
                      FVRHHSRGKAIMTITNQTRYHGYDVKSNEEVKNQALFALQKMMLNNWEYLNAK"
     CDS              complement(13818. .14342)
                      /note="ORF_ID:dd_02721"
                      /codon_start=1
                      /product="hypothetical protein"
                      /protein_id="AAM44366.1"
                      /db_xref="GI:21240655"
                      /translation="MNENFKIIDKEITEILLKNADFIHHIFKIHYKSLIHNNTDNNNN
                      DNSGYNEYWNRIIYKDCNNISIPMFIKFKHLFIGNDCLKPINDKDLELLFGQYCFDID
                      LRNLIINSIPNSNYFNITNSMISKAIDSNRIDYLKYLYENGGEIKNNPSLLQHLKNQL
                      SKSYDYNCFLDWLN"
     CDS              complement(14672. .17431)
                      /note="ORF_ID:dd_02711"
                      /codon_start=1
                      /product="hypothetical protein"
                      /protein_id="AAM44367.1"
                      /db_xref="GI:21240656"
                      /translation="MESPNVEILFWKVFHNKYLFSIILKEIHKCEWVNYVYLGYHVDN
                      RKRFENITSMEWMINNEQFELLKLKINSNDEIIEINKKSIKELFSINDPSLIEQLLTQ
                      LFQSNKKNQFIKIINEMLPTLPIITSNSLRFKNTYNIFQLLLSNLSSLSSSFQIYSTL
                      LECAIEFLQPNTLSILINSPNLFITQDFKNRSLTFTNLNSTSIKTIEEKIELILSNPK
                      LYNTDYSNLDININFNGNGRERILKFIPKNENFGLSLILLINNSNIQNKLIDLNFHEL
                      NLTNLVDNKYKINEILLDYNRELRVKELHKYLLLNPTIVNLYDSIYLDHLVYYSRYGE
                      IIPNLKENNLMDFIKISIATFNVKALKTISKQDLEQFKLKYSKPHSRETTIPLSISRE
                      YKSKQEMIEFINFYTTNYDYFQCFHFNHNKIIQDYILKNLSKTSSSSSSNIEIEYKEL
                      LNSINNIEFLNFFPNSWNELIKIIADSNSILLSVEIKDNLQIDWIFENCEKEFLKRSI
                      RNFKSQEFFHFQNFQLLNYASNLFSSLPRDHYLYFNNNNNQDDDDYTQINSRMVLYSY
                      PFIKSIRDCNIEQMKYLYSTKGDDFFLYGFNITIRLFAEKQFFFSSNLSDEEIINFIQ
                      YLNDNQFKFFKPNLVKRVIKSIPDRIKENELLKLKINFNDFIDEASHIFLKSIEKLSI
                      NQIKLMIGNGFFPIQDISSSLLIRYKDPIYSPLSKCAKFYSISNDNCEKVITLIQIFL
                      QMYENYTKEEDDDNDNDQDYAFHIIINHLFKVLLSVKKVTIQVIKNIYNQISSSGSSS
                      SKHIIIDHSLYHVVFYLSIKSPLLARYIMSRPFIYSIFTDDDDGTTPYYNCSDNPQHP
                      GYDIRKYLNLHVFTFDFILGKDTDTHTSFDLILEKLIDALSYKPMTDDESLLFTEEDY
                      LNETD"
     CDS              18595. .20304
                      /note="ORF_ID:dd_02710"
                      /codon_start=1
                      /product="PAXILLIN-LIKE PROTEIN"
                      /protein_id="AAM44368.1"
                      /db_xref="GI:21240657"
                      /translation="MATKGLNMDDLDLLLADLGRPKSSIKVTATVQTTATPSSGKNFD
                      NSDIQNEIQSIIEELDQQPQTVQTISTPAPKNHNTTTTTASFSVSSQPAPQPPQQSQQ
                      IDGLDDLDELMESLNTSISTALKAVPTTPEEHITHANSNSPPPSLHKNTSSTNSASSL
                      SRPNNNPSVVSTPQPGKVTSTATITTKKQPALSKATLETTSGNNVYSSQPSQSQPQPY
                      KVTATNSQPSSDDLDELLKGLSPSTTTTTTVPPPVQRDQHQHHHQHQHHHHHNPNHNQ
                      TQTVTTQINIGRTNTPNNNNNNNTNSPKVVHGDDLDNLLNNLTSQVKDIDSTGPTSRG
                      TCGGCRKPIFGETIQAMGKFYHPEHFCCHNCQNPLGTKNYYEQESLPHCEKCYQELFC
                      ARCAHCDEPISDRCITALGKKWHVHHFVCTQCLKPFEGGNFFERDGRPYCEADFYSTF
                      AVRCGGCNSPIRGECINALGTQWHPEHFVCQYCQKSFTNGQFFEFGGKPYCDVHYHQQ
                      AGSVCSGCGKAVSGRCVDALDKKWHPEHFVCAFCMNPLAGGSYTANNGKPYCKGCHNK
                      LFA"
     CDS              join(21782. .22563,22679. .22793,22936. .27681)
                      /note="ORF_ID:dd_03308"
                      /codon_start=1
                      /product="putative protein contains similarity to
                      topoisomerase-related function protein"
```

```
                     /protein_id="AAM44369.1"
                     /db_xref="GI:21240658"
                     /translation="MDFKKQVLSWVLSIGHEELETILTDCNKDECDKLRAMYSEWKFQ
                     DGYFFDESVDIPGMPSFDSDSIDDENNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
                     NNNIGDNNNISNNNINNNNNNINNNNNTINNNNRKLATSFSTSNIINNNSTNNINNSS
                     SSSSNNNNNSINGSNTPNNTSINSNCSSNSSSKLKSKKSHSSGSIGNVVCYSDIEMD
                     RRACSSLLKKFKLEFRLTNDKSIYHFENLQQFLPERKVEEEKRIVKLIRLCDTRTRLD
                     TFTIESNSVSNQELVHLLTELSQGGFLQVPCHLIKDSQNNWRSEPPKWFISKSFFSLG
                     EFVAAWFESLIWAQFWENHRIDGRKDQSSKQRNPKPIFKLLDSRKSNLKEFWSKQMTI
                     DNRKRYAMAKMGNLLLSVIHSLNETPEKVDLDFIEKLGELIEDFKSPTSTSPSIISSN
                     TTTPTSTTPTTSISTTTTTTTTSSNNNSNNNNNNNNISSSTPTTPITVSVDKPKLIQ
                     RSTTNRNTNNQSQQNHVSVGGKLGVLGGSSSLSESPENLNSISSIFEEYLVQNPIKLI
                     EFLCFSPLERAGSTVDTVMRQIGAQLQNFYSESLSMDLINSEESSRNRKRIGGNNSNS
                     GSLNSSGNNTPVKQKKKKKQQDQKQQDKEKEKEKEKEKEKEKEKTISPTTTTTTTTTT
                     ANSNSLSNSGGGNSSKKLNESSDSTQSDSTNRTKPNANPNTNTSIRKKRKKKSGRGGN
                     NNNNNNNNNQGRGGNSNSNHGGHHHHNHHQSNQGEKSSANSDTEQESIQTRPRRHSLP
                     MDTPPTIKNGIPFLSTTEDEDGVSMLSESDIDHGKGGTGPIKNNNNNNNNNNNNNNNN
                     NNNNNNNNNNNNNNNNNNNNNNSNHSSNKSRNISTSSSTSTSTSTNNTNNNSNSSTIG
                     GNSLSSSTTGTSQPLSNITTTKPKQNGKTGTNKKGNSKSSSSTPTKPTVTKNRANSTC
                     EQRTPLILTQKSLTSKSLNNIAVTSTSDSDGIKPKFIYSEEFEEDDEDEGSDEESDED
                     DEDEESDEEYREEDGFNDLEDDEEDQNNSYVQDLTKQIEKHFSLNDEKKNNNNNNNSQ
                     DETLKQPIAKQHQQQQQNQQQQQQVNSSSKLNLNQSLDNIPNANNNNQNGQKVLSPQQ
                     TSQQSNTSPRLFNIFATPNSPNPSTSPILLHQATINSSPKVSSPQPLSPKNGSLSPAG
                     NVSSKRLFFENQTISSGNTSPKITNNNNNSNNINNNNSTTTTTTTTNNNNPIATNIQS
                     NINNNEKDKERDNNQKNNNNNNNDNNNNNNNNNINNNNNNTSNNNNNNNNNINNNNN
                     ISNNNNNNNNNNNNNNNNNNNINNNNNTMNSSESTNTNTNTTTTTTTTTTNINNNDR

Alignment Scores:
Pred. No.:              1.23e+04        Length:         61052
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AC123513 (1-61052)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      19226 CACAATCACAACCACAAC 19243

RESULT 461
AC023446/c
LOCUS       AC023446               63087 bp    DNA     linear   HTG 13-JUL-2000
DEFINITION  Homo sapiens clone RP11-817M3, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC023446
VERSION     AC023446.2  GI:9161828
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 63087)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-817M3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 63087)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
            Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
            McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
            Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
            Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 13, 2000 this sequence version replaced gi:6970666.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L6384
                Center clone name: 817_M_3
            --------------.
            * NOTE: This record contains 77 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      734: contig of 734 bp in length
            *      735      834: gap of 100 bp
            *      835     1571: contig of 737 bp in length
            *     1572     1671: gap of 100 bp
            *     1672     2391: contig of 720 bp in length
            *     2392     2491: gap of 100 bp
            *     2492     3227: contig of 736 bp in length
            *     3228     3327: gap of 100 bp
            *     3328     4056: contig of 729 bp in length
            *     4057     4156: gap of 100 bp
            *     4157     4875: contig of 719 bp in length
            *     4876     4975: gap of 100 bp
            *     4976     5695: contig of 720 bp in length
            *     5696     5795: gap of 100 bp
            *     5796     6541: contig of 746 bp in length
            *     6542     6641: gap of 100 bp
            *     6642     7371: contig of 730 bp in length
            *     7372     7471: gap of 100 bp
            *     7472     8183: contig of 712 bp in length
            *     8184     8283: gap of 100 bp
            *     8284     9025: contig of 742 bp in length
            *     9026     9125: gap of 100 bp
            *     9126     9849: contig of 724 bp in length
            *     9850     9949: gap of 100 bp
            *     9950    10685: contig of 736 bp in length
            *    10686    10785: gap of 100 bp
            *    10786    11516: contig of 731 bp in length
            *    11517    11616: gap of 100 bp
            *    11617    12310: contig of 694 bp in length
            *    12311    12410: gap of 100 bp
            *    12411    13124: contig of 714 bp in length
            *    13125    13224: gap of 100 bp
            *    13225    13940: contig of 716 bp in length
            *    13941    14040: gap of 100 bp
            *    14041    14757: contig of 717 bp in length
            *    14758    14857: gap of 100 bp
            *    14858    15582: contig of 725 bp in length
            *    15583    15682: gap of 100 bp
            *    15683    16404: contig of 722 bp in length
            *    16405    16504: gap of 100 bp
```

```
*    16505    17234: contig of 730 bp in length
*    17235    17334: gap of 100 bp
*    17335    18053: contig of 719 bp in length
*    18054    18153: gap of 100 bp
*    18154    18884: contig of 731 bp in length
*    18885    18984: gap of 100 bp
*    18985    19716: contig of 732 bp in length
*    19717    19816: gap of 100 bp
*    19817    20538: contig of 722 bp in length
*    20539    20638: gap of 100 bp
*    20639    21361: contig of 723 bp in length
*    21362    21461: gap of 100 bp
*    21462    22186: contig of 725 bp in length
*    22187    22286: gap of 100 bp
*    22287    23012: contig of 726 bp in length
*    23013    23112: gap of 100 bp
*    23113    23839: contig of 727 bp in length
*    23840    23939: gap of 100 bp
*    23940    24670: contig of 731 bp in length
*    24671    24770: gap of 100 bp
*    24771    25494: contig of 724 bp in length
*    25495    25594: gap of 100 bp
*    25595    26311: contig of 717 bp in length
*    26312    26411: gap of 100 bp
*    26412    27134: contig of 723 bp in length
*    27135    27234: gap of 100 bp
*    27235    27938: contig of 704 bp in length
*    27939    28038: gap of 100 bp
*    28039    28733: contig of 695 bp in length
*    28734    28833: gap of 100 bp
*    28834    29549: contig of 716 bp in length
*    29550    29649: gap of 100 bp
*    29650    30378: contig of 729 bp in length
*    30379    30478: gap of 100 bp
*    30479    31210: contig of 732 bp in length
*    31211    31310: gap of 100 bp
*    31311    31948: contig of 638 bp in length
*    31949    32048: gap of 100 bp
*    32049    32771: contig of 723 bp in length
*    32772    32871: gap of 100 bp
*    32872    33609: contig of 738 bp in length
*    33610    33709: gap of 100 bp
*    33710    34432: contig of 723 bp in length
*    34433    34532: gap of 100 bp
*    34533    35259: contig of 727 bp in length
*    35260    35359: gap of 100 bp
*    35360    36079: contig of 720 bp in length
*    36080    36179: gap of 100 bp
*    36180    36871: contig of 692 bp in length
*    36872    36971: gap of 100 bp
*    36972    37698: contig of 727 bp in length
*    37699    37798: gap of 100 bp
*    37799    38530: contig of 732 bp in length
*    38531    38630: gap of 100 bp
*    38631    39346: contig of 716 bp in length
*    39347    39446: gap of 100 bp
*    39447    40175: contig of 729 bp in length
*    40176    40275: gap of 100 bp
*    40276    40977: contig of 702 bp in length
*    40978    41077: gap of 100 bp
*    41078    41804: contig of 727 bp in length
*    41805    41904: gap of 100 bp
*    41905    42623: contig of 719 bp in length
*    42624    42723: gap of 100 bp
*    42724    43417: contig of 694 bp in length
*    43418    43517: gap of 100 bp
*    43518    44245: contig of 728 bp in length
*    44246    44345: gap of 100 bp
*    44346    45064: contig of 719 bp in length
*    45065    45164: gap of 100 bp
*    45165    45856: contig of 692 bp in length
*    45857    45956: gap of 100 bp
*    45957    46680: contig of 724 bp in length
*    46681    46780: gap of 100 bp
*    46781    47513: contig of 733 bp in length
*    47514    47613: gap of 100 bp
*    47614    48338: contig of 725 bp in length
*    48339    48438: gap of 100 bp
*    48439    49157: contig of 719 bp in length
*    49158    49257: gap of 100 bp
*    49258    49968: contig of 711 bp in length
*    49969    50068: gap of 100 bp
*    50069    50778: contig of 710 bp in length
*    50779    50878: gap of 100 bp
*    50879    51594: contig of 716 bp in length
*    51595    51694: gap of 100 bp
*    51695    52416: contig of 722 bp in length
*    52417    52516: gap of 100 bp
*    52517    53248: contig of 732 bp in length
*    53249    53348: gap of 100 bp
*    53349    54065: contig of 717 bp in length
*    54066    54165: gap of 100 bp
*    54166    54894: contig of 729 bp in length
*    54895    54994: gap of 100 bp
*    54995    55720: contig of 726 bp in length
*    55721    55820: gap of 100 bp
*    55821    56521: contig of 701 bp in length

Alignment Scores:
Pred. No.:                1.27e+04        Length:          63087
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       14              Gaps:            0

US-09-858-332G-28 (1-6) x AC023446 (1-63087)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      13368 CATAATCATAATCATAAT 13351

RESULT 462
AL833790
LOCUS       AL833790               63139 bp    DNA     linear   ROD 29-AUG-2002
DEFINITION  Mouse DNA sequence from clone RP23-24E16 on chromosome X, complete
            sequence.
ACCESSION   AL833790
VERSION     AL833790.6  GI:22552878
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 63139)
  AUTHORS   Howden,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 30, 2002 this sequence version replaced gi:22416192.
            ------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
```

```
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-24E16 is
            from the RPCI-23 Mouse PAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES                Location/Qualifiers
     source             1. .63139
                        /organism="Mus musculus"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:10090"
                        /chromosome="X"
                        /clone="RP23-24E16"
                        /clone_lib="RPCI-23"
ORIGIN

Alignment Scores:
Pred. No.:              1.27e+04        Length:         63139
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     9               Gaps:           0

US-09-858-332G-28 (1-6) x AL833790 (1-63139)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      16869 CATAATCATAACCATAAT 16886

RESULT 463
CQ600515/c
LOCUS       CQ600515                63284 bp    DNA     linear   PAT 02-FEB-2004
DEFINITION  Sequence 28273 from Patent WO0171042.
ACCESSION   CQ600515
VERSION     CQ600515.1  GI:41655656
KEYWORDS    .
SOURCE      Drosophila sp.
  ORGANISM  Drosophila sp.
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1
  AUTHORS   Venter,J.C., Adams,M., Li,P.W. and Myers,E.W.
  TITLE     Detection kits, such as nucleic acid arrays, for detecting the
            expression of 10,000 or more Drosophila genes and uses thereof
  JOURNAL   Patent: WO 0171042-A 28273 27-SEP-2001;
            PE Corporation (NY) (US)
FEATURES                Location/Qualifiers
     source             1. .63284
                        /organism="Drosophila sp."
                        /mol_type="unassigned DNA"
                        /db_xref="taxon:7242"
ORIGIN

Alignment Scores:
Pred. No.:              1.27e+04        Length:         63284
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x CQ600515 (1-63284)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      13155 CATAATCATAATCATAAT 13138

RESULT 464
H0410G08
LOCUS       H0410G08                63554 bp    DNA     linear   PLN 08-JAN-2001
DEFINITION  Oryza sativa genomic DNA, chromosome4, BAC clone: H0410G08,
            complete sequence.
ACCESSION   AL512546
VERSION     AL512546.1  GI:12140342
KEYWORDS    .
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Han,B., Zhou,B., Feng,Q., Mu,J., Chen,Z.H., Li,Y., Zhu,J.J.,
            Tang,Y.S., Zhao,Q., Liu,Y.L., Huang,Y.C., Yu,Z., Fan,D.L., Chen,L.,
            Weng,Q.J., Zhang,L., Lu,Y.Q., Yu,S.L., Zhu,J., Liu,X.H., Hu,X.,
            Lei,H.Y., Zhang,Y.J., Wang,R., Li,C., Lu,Y., Chen,X.C., Zhang,Y.,
            Hu,H., Jia,P.X., Li,T., Qian,Y.M., Ying,K. and Hong,G.F.
  TITLE     Oryza sativa indica (Guangluai4) genomic DNA, chromosome 4, BAC
            clone: H0410G08
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 63554)
  AUTHORS   Hong,G.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JAN-2001) Hong G.F., National Center for Gene
            Research, Chinese Academy of Sciences, 500# Caobao Road, Shanghai
            200233, CHINA
FEATURES                Location/Qualifiers
     source             1. .63554
                        /organism="Oryza sativa (indica cultivar-group)"
                        /mol_type="genomic DNA"
                        /variety="GuangLuAi4"
                        /sub_species="indica"
                        /db_xref="taxon:39946"
                        /chromosome="4"
                        /clone="H0410G08"
                        /clone_lib="BAC-HindIII"
ORIGIN

Alignment Scores:
Pred. No.:              1.28e+04        Length:         63554
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x H0410G08 (1-63554)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      47678 CACAACCACAACCACAAC 47695

RESULT 465
AP002455
LOCUS       AP002455                64000 bp    DNA     linear   PRI 15-MAR-2003
DEFINITION  Homo sapiens genomic DNA, chromosome 11q, clone:RP11-93M11,
            complete sequence.
ACCESSION   AP002455
VERSION     AP002455.4  GI:17425240
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
```

```
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA
  JOURNAL   Published Only in Database (2000)
REFERENCE   2  (bases 1 to 64000)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     On Dec 7, 2001 this sequence version replaced gi:11320829.
FEATURES             Location/Qualifiers
     source          1. .64000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-93M11"
ORIGIN

Alignment Scores:
Pred. No.:                1.28e+04        Length:         64000
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       8               Gaps:           0

US-09-858-332G-28 (1-6) x AP002455 (1-64000)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      59121 CATAACCACAATCATAAT 59138

RESULT 466
AL445646/c
LOCUS       AL445646               64120 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP11-217F16 on chromosome 9 Contains
            a pseudogene similar to part of member RAB7 of RAS oncogene family
            (RAB7) and part of a ribosomal protein S12 (RPS12) pseudogene,
            complete sequence.
ACCESSION   AL445646
VERSION     AL445646.11  GI:14160979
KEYWORDS    HTG; RAB7; ras; RPS12.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 64120)
  AUTHORS   Kimberley,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 20, 2001 this sequence version replaced gi:13375013.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            RP11-217F16 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1. .64120
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-217F16"
                     /clone_lib="RPCI-11.1"
     misc_feature    100
                     /note="Clone_right_end: RP11-63P12"
     gene            complement(39411. .39614)
                     /locus_tag="RP11-217F16.1-001"
                     /pseudo
     CDS             complement(39411. .39614)
                     /locus_tag="RP11-217F16.1-001"
                     /note="match: proteins: AAH04597 AAM21090 BAB23738 P09527
                     P18067 P51149 P51150"
                     /pseudo
                     /codon_start=1
     misc_feature    53757
                     /note="Alu region"
     misc_feature    53774
                     /note="Alu region."
     misc_feature    53784
                     /note="Alu region"
     gene            join(64036. .64120,AL596448.5:2001. .2155,
                     AL596448.5:2467. .2618)
                     /locus_tag="RP11-313E4.1-001"
                     /pseudo
     CDS             join(64036. .64120,AL596448.5:2001. .2155,
                     AL596448.5:2467. .2618)
                     /locus_tag="RP11-313E4.1-001"
                     /note="match: proteins: AAH17321 AAH18362 BAB22404
                     BAB79478 P09388 P25398 P46405 P47840"
                     /pseudo
                     /codon_start=1
     misc_feature    64120
                     /locus_tag="RP11-313E4.1-001"
                     /note="Clone_right_end: RP11-217F16"
                     /pseudo
ORIGIN

Alignment Scores:
Pred. No.:                1.29e+04        Length:         64120
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       8               Gaps:           0

US-09-858-332G-28 (1-6) x AL445646 (1-64120)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      17066 CATAACCATAATCATAAT 17049

RESULT 467
CR388210
LOCUS       CR388210               64196 bp    DNA     linear   PRI 29-OCT-2004
DEFINITION  Human DNA sequence from clone DAMA-186H17 on chromosome 6, complete
            sequence.
```

```
ACCESSION   CR388210
VERSION     CR388210.7  GI:54888682
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 64196)
  AUTHORS   Wood,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 29, 2004 this sequence version replaced gi:54633111.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs constructed by
            the MHC Haplotype Consortium and collaborators. Further information
            can be found at
            http://www.sanger.ac.uk/HGP/Chr6/MHC
            DAMA-186H17 is from the DNA-Arts human BAC library MANN.1 VECTOR:
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .64196
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="DAMA-186H17"
                     /clone_lib="DNA-Arts BAC library MANN.1"
ORIGIN

Alignment Scores:
Pred. No.:              1.29e+04        Length:         64196
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     8               Gaps:           0

US-09-858-332G-28 (1-6) x CR388210 (1-64196)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       3640 CATAATCATAATCATAAT 3657

RESULT 468
AC115607/c
LOCUS       AC115607               64707 bp    DNA     linear   INV 21-FEB-2004
DEFINITION  Dictyostelium discoideum chromosome 2 map 4559693-4624405 strain
            AX4, complete sequence.
ACCESSION   AC115607 AC116959
VERSION     AC115607.2  GI:28829478
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 64707)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 64707)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 64707)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 64707)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   5  (bases 1 to 64707)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2004) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On or before Feb 21, 2004 this sequence version replaced
            gi:28850456, gi:19570077.
            CDS predictions from GeneID may contain errors. Further Information
            is available from IMB Jena, Department of Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES             Location/Qualifiers
     source          1. .64707
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /chromosome="2"
                     /map="4559693-4624405"
     CDS             1336. .4150
                     /note="GeneID exon scores (in order of location ranges):
                     199.30 - GSCJ_ID dd_01223"
                     /pseudo
                     /codon_start=1
     CDS             complement(join(4548. .4688,4778. .4889,4959. .5230,
                     5391. .5624,6033. .6154,6498. .6564))
                     /note="GeneID exon scores (in order of location ranges):
                     10.05, 6.71, 31.71, 28.25, 17.96, 5.94 - GSCJ_ID dd_01225"
                     /codon_start=1
                     /product="similar to Homo sapiens (Human). Hypothetical
                     protein FLJ11197"
                     /protein_id="AAL92367.2"
                     /db_xref="GI:28829479"
                     /translation="MGQLLSFINGNDHTEQIFIDFEHAQPSDDERELHKTVNEVLIRG
                     PAIIDKLLAYAGCNEFIRRAITNPGPETEDAAWEAVLPSVDQLQEFYDFSLELETCFP
                     KLLVALCKNDPKASLSNQQALAKQLADIFDFVLKFDDAKMVNPAIQNDFSYYRRTLNR
                     MKLTKKDANIKIRDELANRMSLFPAYPTPMMKVLSETTVKFLSQDTTVPRDNVTTALA
                     TMANVCHDMVEKKKFTQDDLNMFCLRAMVGSIILFDHIHPQGAFVKKSPVNIKPCIVT
```

LKDSNTQSSPGLLNALRFTTIHLNDVDTPGAIKQLLLLD"
CDS 7749. .9245
/note="GeneID exon scores (in order of location ranges): 107.75 - GSCJ_ID dd_01227"
/codon_start=1
/product="similar to Homo sapiens (Human). similar to RIKEN cDNA 2810423E13 gene"
/protein_id="AAL92368.2"
/db_xref="GI:28829480"
/translation="MSYKIINSSDDDNDNNNNNDINNNENIIREEEKNKTLPNSPIQT
DEQKKKQTLSINIMVLTVFLQSIGFTLILPSMLNYLNSNDPQFKRYFGYIVALYSLGQ
FLGSPLFGKWSNKRAASEPVVISIVISIIGSIFYSICYKFHGFAFPAIMGTARFIVGF
GAGNVSVCRAYATETATIENKTQIMGKMSGAQGAGFVLGPGIGFLLNFCNFTIGQMVI
NKYTAPGYLSILLAILNIILVLVSFRDARDLEKRNLKTPETASLLSINDADADADANN
DADKPKKNMEQRLKPIVLSIFLFMVVITIFAVFETVLTLMTLQYYDWGSTENYIILGT
SGIISVGIFIAISSPVIKKFDDRKTALFGFTCLFIALVLLINYNIPIGSRMTLPKWQF
FMGSAFVSVGYPIASSLVYAIFSKVLNPNSHTQGTKMGWLTAGGSLARMVGPLWSQEI
WTATNQQGMELFFVTAVITLIAIVVFIIFYKDLSPHPEFVSQQKTVNK"
CDS join(11117. .11358,11490. .12591)
/note="GeneID exon scores (in order of location ranges): 13.04, 88.43 - GSCJ_ID dd_01228"
/codon_start=1
/product="similar to Homo sapiens (Human). Testis intercellular mediator Peas (Sortilin 1) (Hypothetical protein) (Testis intracellular mediator protein)"
/protein_id="AAL92369.2"
/db_xref="GI:28829481"
/translation="MTNNSNIENCNTIYDKQITQFKINRNGSSGCGSSNESGSLLISI
KIEHILEFPDEVLLTILSFLEPNDIKNIYLSSRYLSRFCFENKIWKAICKRKWSSSPI
FKRKPITNWRSYYLKKKGFMDLSGGLSWIEIFSSSGDKPSPRYQNTCTSVGNEIYFIG
GQDHPENRFNTIHKFNCTTQEFTRLIPTGSSPPKFARHTSTVIGSKIYSHGGFDGFSQ
HFGLFVYCTIENRWETLSPTGDIPISRTNHASTSIGDNFYIFGGMYKNGSDLVFLNDL
FEYNTIENKWTQLKGTGDHPPQKCGHKLFNFGGKLLLFGGGYGSQWDIKYNDVHIYDR
SLNRWTKVKTKGEINVCTFTVSWSVGPFLFIYGGQSVKDDLLTNDLYMLDTVNMEWSS
INTNCTPFPRDMGSGVLVGSTLYMFGGFSSTPLDSLFSLDLLKDENLNISPSPNI"
CDS complement(12707. .14728)
/note="GeneID exon scores (in order of location ranges): 186.10 - GSCJ_ID dd_01229"
/codon_start=1
/product="similar to Dictyostelium discoideum (Slime mold). TRFA"
/protein_id="AAL92370.2"
/db_xref="GI:28829482"
/translation="MKNYNNNTAHLKALQAEISKKKFIRRYLCKKIVEDRIEDQYNKK
EQEQKEIEQKLKLEQEQKEIELKKQQQSLLTEEQRLQQEQQEMEQRQKTITLKREKLQ
KKVNSLKDLQQQQQSKIIVQKEILETLESNKNTLFLLLKTVMAEERRQTEIQKALFLK
KQQQQQQQFLYQQQQQQQQQQQQQQQQQQQQQQHDQQQQQHDQQQDQQKQHDQQQHQEQ
QPQQQQFNNNPNFNGNTTNNSNQFMNGQNIQFNNNDIHQSQPQPQSQPQPQPQPQPQP
QPQPQPQPQPQPQPQQPQPQPQPQQQQQLQFSSNNNGTFNNTNNYNNGSFNNNNNNNN
NNNNNNSGFINSSNGNNFNSYNNNSGNSNNNNNNNNNNNVSQYSNGVNGNNNNYNNGF
NNNINGSIDNNNNNNTTNSNNVYNYQNKSYSVDDRSSYNHKPNFMQPSSPSMQGQYYP
STKPITGMSPLSPFSSPVSPSQTLFINNSNSNSNVNSFNNNNNNNNSSNNSNNSNSN
RISPNLTSSNMYRNDKDQFYNSNNSNKGNYNNANNSNNNNNRFYEKDRDQYNNNNIN
NNNNNNNNSNNNNNNNNNNNNNNNNNNNNNINNNNNNNGNGYHYQNYNNNDNRGYNNNRR
SDNDNSYEDGSSNNNNYNNNNRGHPYKNNYHHNNNRRIYNNNNWKSQTN"
CDS 16507. .16674
/note="GeneID exon scores (in order of location ranges): 3.53 - GSCJ_ID dd_01220"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAO52016.1"
/db_xref="GI:28829483"
/translation="MEKIKKQNNIGSYSSDKDGEYKKFRFSKTSIRGFLLKEYEREVP
IIVKIQVMYII"
CDS join(16869. .17233,17321. .17990)
/note="GeneID exon scores (in order of location ranges): 28.66, 68.01 - GSCJ_ID dd_01221"
/codon_start=1
/product="similar to putative sphingoid base-phosphate phosphatase [Schizosaccharomyces pombe]"
/protein_id="AAO52017.1"
/db_xref="GI:28829484"
/translation="MASILGEEVFFILALPISTWCVATQLGVELCVVLALTIGGGNIL
KNTFTLPRPPPNIVWTNTAHQKDHGLPSTHTASAFGLTFYFLIYTYFLFPTIGESFNI
SLLSMFFIVLFWSTSVMFSRLYNGHHTPMDVIAGLIVAITSILATTYQLRYFIEGMVL
SETFLFGPMLYIAILSAILFFHPQANTGPTPAYPETGLVCGASLGSLISLWLHAQHPC
PLMNQELLLLEANYDSIVSTIHSIPLLLHGSRILIGLVLVGIAKVFSKKFFFFAYDLV
IRANTNNEQSQPITTVSFDPNKKIIVTPTIEAFSKLFVYTCVSFTIVSMPYLFYYLNI
QTSADVTRYY"
CDS join(18447. .18472,18540. .18883,19090. .19319,19444. .19559,19677. .21864)
/note="GeneID exon scores (in order of location ranges): -0.96, 4.97, 2.47, 5.06, 92.96 - GSCJ_ID dd_01222"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAO52018.1"
/db_xref="GI:28829485"
/translation="MKIYQPFINNSITVFFQRNSKHDFLLMVYYSLLFTSSFCIGSVL
LVISSIRHFFLAFELFWLLVIATTAIVSSEHFYGIIKNVIRLCVAVIIGSILSYISFS
MVGRSVWPNCFIFFVILLLGTETFRKEFYSPIDNLLFTLGSAASFGSCLLTCLLLRAP
FANILFIQTSNTLLKRSISYIKQLKLQTNLVFSRNNNFKIYYRDIDENSKKKTNDLIL
KLQFQHQDYSERLNHLDQMLSDSKKEFWAKNYIQHFERIQILFERNLKKLIAVGLKIN
DEIDDKIRNQIEPIIPQVNILFDEILKIFEEMGNQLKYKFIHTNNNDNSQWFKNTIAS
DYYKNLDKRDENYHSFFYCHLDFVYKINGLKDIKENSLEWNQIKINKSFEIIDETISK
IDQLFESIFKEAPLFNPNVDNSLSKVYILLKCFNTFCLEQKLLSKEIFLVSYHNSLNR
NKPQEYLAIVATIKRIWKYTKTYKQRKLEKLENERNTNEDTTPIKEKFKLYFMNFLKD
RILYNWQFTIKYSLVSAGGAVAIFEIKKHTKFQLFEMLNWMLNAYIVTAGPDIGALGA
LTSVRVIVLIIGGFLGYASIIICEFGGNDSSKALIYIGFSFAIIFLLGIILNFQIFKG
MITNIVFTFSVVSIPLYENGSNIIFTLYRIGYVVLGFILVFILSMIIPYYDYRELEKN
LFKIPFLIIELVDYFFYYSFEQPIHESNKTLLSIPYNENNENYVFNNIDYNNFSHVNN
INSEKSFKISEMITIDQYKSVFSEKYINLRRIFPAQRNLLLDSKLELIFKKSEFKNLE
NDLLSIQYFHDLFGVLEFAFVESHETHIRMIGKETRIQARRLFDELELSSTLLCHIQE
ERDNLGRVIRSSSLPYISQKRDNGIDLMNQLIDHVNSQISNYNIPTLEKHQLNATLFA
LHSFVKQYHLIFKKLILKKESMVNNSKVFILNNNNYKINSNMIDVKESIFLNQ"
CDS complement(join(22090. .22485,22601. .22978,23168. .23311))
/note="GeneID exon scores (in order of location ranges): 42.25, 39.59, 19.88 - GSCJ_ID dd_01224"
/codon_start=1
/product="similar to Dictyostelium discoideum (Slime mold). Protein phosphatase 4 catalytic subunit (EC 3.1.3.16) (Serine/threonine protein phosphatase)"
/protein_id="AAO52019.1"
/db_xref="GI:28829486"
/translation="MSSDLDRQIEQLKRCEIIKESEVRALCSKAREILLEEGNVQRVD
SPVTICGDIHGQFYDLKELFKVGGDCPQTNYLFMGDFVDRGFYSVETFLLLLALKVRY
PDRITLIRGNHESRQITQVYGFYEECVRKYGSVTVWKYCTEIFDYLSLSALVDGKIFC
VHGGLSPSINTLDQIRAIDRKQEVPHEGPMCDLMWSDPEDIPGWNGSPRGAGFLFGED
VVQKFNHDNNLEFICRAHQLVMEGFKYMFNETLVTVWSAPNYCYRCGNVAAILQLDEN
LKKNFAIFEAAPQESRGAPAKKPAPEYFL"
CDS 24836. .25615
/note="GeneID exon scores (in order of location ranges): 49.65 - GSCJ_ID dd_03410"
/codon_start=1
/product="similar to Plasmodium falciparum. Hypothetical protein"
/protein_id="AAO52020.1"

Alignment Scores:

Pred. No.:	1.3e+04	Length:	64707
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-28 (1-6) x AC115607 (1-64707)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      13956 CACAACCACAACCACAAT 13939
```

RESULT 469
AC090905/c
LOCUS AC090905 64958 bp DNA linear HTG 17-MAR-2001
DEFINITION Homo sapiens chromosome 15 clone RP11-327C2 map 15, LOW-PASS SEQUENCE SAMPLING.

```
ACCESSION   AC090905
VERSION     AC090905.1  GI:13374683
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 64958)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-327C2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 64958)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
            Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K.,
            Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
            Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L13128
                Center clone name: 327_C_2
            --------------.
            * NOTE: This record contains 82 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      696: contig of 696 bp in length
            *      697      796: gap of 100 bp
            *      797     1492: contig of 696 bp in length
            *     1493     1592: gap of 100 bp
            *     1593     2293: contig of 701 bp in length
            *     2294     2393: gap of 100 bp
            *     2394     3082: contig of 689 bp in length
            *     3083     3182: gap of 100 bp
            *     3183     3872: contig of 690 bp in length
            *     3873     3972: gap of 100 bp
            *     3973     4672: contig of 700 bp in length
            *     4673     4772: gap of 100 bp
            *     4773     5461: contig of 689 bp in length
            *     5462     5561: gap of 100 bp
            *     5562     6256: contig of 695 bp in length
            *     6257     6356: gap of 100 bp
            *     6357     7044: contig of 688 bp in length
            *     7045     7144: gap of 100 bp
            *     7145     7839: contig of 695 bp in length
            *     7840     7939: gap of 100 bp
            *     7940     8623: contig of 684 bp in length
            *     8624     8723: gap of 100 bp
            *     8724     9430: contig of 707 bp in length
            *     9431     9530: gap of 100 bp
            *     9531    10231: contig of 701 bp in length
            *    10232    10331: gap of 100 bp
            *    10332    11018: contig of 687 bp in length
            *    11019    11118: gap of 100 bp
            *    11119    11829: contig of 711 bp in length
            *    11830    11929: gap of 100 bp
            *    11930    12604: contig of 675 bp in length
            *    12605    12704: gap of 100 bp
            *    12705    13405: contig of 701 bp in length
            *    13406    13505: gap of 100 bp
            *    13506    14202: contig of 697 bp in length
            *    14203    14302: gap of 100 bp
            *    14303    14991: contig of 689 bp in length
            *    14992    15091: gap of 100 bp
            *    15092    15788: contig of 697 bp in length
            *    15789    15888: gap of 100 bp
            *    15889    16578: contig of 690 bp in length
            *    16579    16678: gap of 100 bp
            *    16679    17358: contig of 680 bp in length
            *    17359    17458: gap of 100 bp
            *    17459    18147: contig of 689 bp in length
            *    18148    18247: gap of 100 bp
            *    18248    18967: contig of 720 bp in length
            *    18968    19067: gap of 100 bp
            *    19068    19756: contig of 689 bp in length
            *    19757    19856: gap of 100 bp
            *    19857    20545: contig of 689 bp in length
            *    20546    20645: gap of 100 bp
            *    20646    21338: contig of 693 bp in length
            *    21339    21438: gap of 100 bp
            *    21439    22133: contig of 695 bp in length
            *    22134    22233: gap of 100 bp
            *    22234    22933: contig of 700 bp in length
            *    22934    23033: gap of 100 bp
            *    23034    23723: contig of 690 bp in length
            *    23724    23823: gap of 100 bp
            *    23824    24512: contig of 689 bp in length
            *    24513    24612: gap of 100 bp
            *    24613    25302: contig of 690 bp in length
            *    25303    25402: gap of 100 bp
            *    25403    26097: contig of 695 bp in length
            *    26098    26197: gap of 100 bp
            *    26198    26907: contig of 710 bp in length
            *    26908    27007: gap of 100 bp
            *    27008    27700: contig of 693 bp in length
            *    27701    27800: gap of 100 bp
            *    27801    28489: contig of 689 bp in length
            *    28490    28589: gap of 100 bp
            *    28590    29282: contig of 693 bp in length
            *    29283    29382: gap of 100 bp
            *    29383    30083: contig of 701 bp in length
            *    30084    30183: gap of 100 bp
            *    30184    30881: contig of 698 bp in length
            *    30882    30981: gap of 100 bp
            *    30982    31676: contig of 695 bp in length
            *    31677    31776: gap of 100 bp
            *    31777    32470: contig of 694 bp in length
            *    32471    32570: gap of 100 bp
            *    32571    33254: contig of 684 bp in length
            *    33255    33354: gap of 100 bp
            *    33355    34043: contig of 689 bp in length
            *    34044    34143: gap of 100 bp
```

```
*    34144    34833: contig of 690 bp in length
*    34834    34933: gap of 100 bp
*    34934    35646: contig of 713 bp in length
*    35647    35746: gap of 100 bp
*    35747    36436: contig of 690 bp in length
*    36437    36536: gap of 100 bp
*    36537    37228: contig of 692 bp in length
*    37229    37328: gap of 100 bp
*    37329    38047: contig of 719 bp in length
*    38048    38147: gap of 100 bp
*    38148    38843: contig of 696 bp in length
*    38844    38943: gap of 100 bp
*    38944    39634: contig of 691 bp in length
*    39635    39734: gap of 100 bp
*    39735    40431: contig of 697 bp in length
*    40432    40531: gap of 100 bp
*    40532    41215: contig of 684 bp in length
*    41216    41315: gap of 100 bp
*    41316    42009: contig of 694 bp in length
*    42010    42109: gap of 100 bp
*    42110    42800: contig of 691 bp in length
*    42801    42900: gap of 100 bp
*    42901    43581: contig of 681 bp in length
*    43582    43681: gap of 100 bp
*    43682    44351: contig of 670 bp in length
*    44352    44451: gap of 100 bp
*    44452    45140: contig of 689 bp in length
*    45141    45240: gap of 100 bp
*    45241    45914: contig of 674 bp in length
*    45915    46014: gap of 100 bp
*    46015    46693: contig of 679 bp in length
*    46694    46793: gap of 100 bp
*    46794    47482: contig of 689 bp in length
*    47483    47582: gap of 100 bp
*    47583    48275: contig of 693 bp in length
*    48276    48375: gap of 100 bp
*    48376    49066: contig of 691 bp in length
*    49067    49166: gap of 100 bp
*    49167    49869: contig of 703 bp in length
*    49870    49969: gap of 100 bp
*    49970    50669: contig of 700 bp in length
*    50670    50769: gap of 100 bp
*    50770    51459: contig of 690 bp in length
*    51460    51559: gap of 100 bp
*    51560    52257: contig of 698 bp in length
*    52258    52357: gap of 100 bp
*    52358    53057: contig of 700 bp in length
*    53058    53157: gap of 100 bp
*    53158    53855: contig of 698 bp in length
*    53856    53955: gap of 100 bp
*    53956    54649: contig of 694 bp in length

Alignment Scores:
Pred. No.:                1.3e+04        Length:          64958
Score:                    42.00          Matches:         6
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       14             Gaps:            0

US-09-858-332G-28 (1-6) x AC090905 (1-64958)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      17814 CATAATCATAATCATAAT 17797

RESULT 470
AC113185
LOCUS       AC113185               65263 bp    DNA     linear   HTG 26-FEB-2002
DEFINITION  Mus musculus clone RP24-195K9, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC113185
VERSION     AC113185.1  GI:18921363
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 65263)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-195K9
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 65263)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-FEB-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L21360
                Center clone name: 195_K_9
            --------------.
            * NOTE: This record contains 81 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      674: contig of 674 bp in length
            *      675      774: gap of 100 bp
            *      775     1480: contig of 706 bp in length
            *     1481     1580: gap of 100 bp
            *     1581     2302: contig of 722 bp in length
            *     2303     2402: gap of 100 bp
            *     2403     3131: contig of 729 bp in length
            *     3132     3231: gap of 100 bp
            *     3232     3872: contig of 641 bp in length
            *     3873     3972: gap of 100 bp
            *     3973     4671: contig of 699 bp in length
            *     4672     4771: gap of 100 bp
            *     4772     5477: contig of 706 bp in length
            *     5478     5577: gap of 100 bp
            *     5578     6284: contig of 707 bp in length
```

```
*     6285     6384: gap of 100 bp
*     6385     7076: contig of 692 bp in length
*     7077     7176: gap of 100 bp
*     7177     7885: contig of 709 bp in length
*     7886     7985: gap of 100 bp
*     7986     8672: contig of 687 bp in length
*     8673     8772: gap of 100 bp
*     8773     9474: contig of 702 bp in length
*     9475     9574: gap of 100 bp
*     9575    10310: contig of 736 bp in length
*    10311    10410: gap of 100 bp
*    10411    11129: contig of 719 bp in length
*    11130    11229: gap of 100 bp
*    11230    11939: contig of 710 bp in length
*    11940    12039: gap of 100 bp
*    12040    12747: contig of 708 bp in length
*    12748    12847: gap of 100 bp
*    12848    13556: contig of 709 bp in length
*    13557    13656: gap of 100 bp
*    13657    14365: contig of 709 bp in length
*    14366    14465: gap of 100 bp
*    14466    15179: contig of 714 bp in length
*    15180    15279: gap of 100 bp
*    15280    15988: contig of 709 bp in length
*    15989    16088: gap of 100 bp
*    16089    16763: contig of 675 bp in length
*    16764    16863: gap of 100 bp
*    16864    17576: contig of 713 bp in length
*    17577    17676: gap of 100 bp
*    17677    18391: contig of 715 bp in length
*    18392    18491: gap of 100 bp
*    18492    19051: contig of 560 bp in length
*    19052    19151: gap of 100 bp
*    19152    19864: contig of 713 bp in length
*    19865    19964: gap of 100 bp
*    19965    20675: contig of 711 bp in length
*    20676    20775: gap of 100 bp
*    20776    21488: contig of 713 bp in length
*    21489    21588: gap of 100 bp
*    21589    22307: contig of 719 bp in length
*    22308    22407: gap of 100 bp
*    22408    23129: contig of 722 bp in length
*    23130    23229: gap of 100 bp
*    23230    23941: contig of 712 bp in length
*    23942    24041: gap of 100 bp
*    24042    24729: contig of 688 bp in length
*    24730    24829: gap of 100 bp
*    24830    25548: contig of 719 bp in length
*    25549    25648: gap of 100 bp
*    25649    26388: contig of 740 bp in length
*    26389    26488: gap of 100 bp
*    26489    27215: contig of 727 bp in length
*    27216    27315: gap of 100 bp
*    27316    28022: contig of 707 bp in length
*    28023    28122: gap of 100 bp
*    28123    28836: contig of 714 bp in length
*    28837    28936: gap of 100 bp
*    28937    29646: contig of 710 bp in length
*    29647    29746: gap of 100 bp
*    29747    30455: contig of 709 bp in length
*    30456    30555: gap of 100 bp
*    30556    31269: contig of 714 bp in length
*    31270    31369: gap of 100 bp
*    31370    32086: contig of 717 bp in length
*    32087    32186: gap of 100 bp
*    32187    32905: contig of 719 bp in length
*    32906    33005: gap of 100 bp
*    33006    33722: contig of 717 bp in length
*    33723    33822: gap of 100 bp
*    33823    34502: contig of 680 bp in length
*    34503    34602: gap of 100 bp
*    34603    35297: contig of 695 bp in length
*    35298    35397: gap of 100 bp
*    35398    36118: contig of 721 bp in length
*    36119    36218: gap of 100 bp
*    36219    36930: contig of 712 bp in length
*    36931    37030: gap of 100 bp
*    37031    37730: contig of 700 bp in length
*    37731    37830: gap of 100 bp
*    37831    38533: contig of 703 bp in length
*    38534    38633: gap of 100 bp
*    38634    39346: contig of 713 bp in length
*    39347    39446: gap of 100 bp
*    39447    40156: contig of 710 bp in length
*    40157    40256: gap of 100 bp
*    40257    40954: contig of 698 bp in length
*    40955    41054: gap of 100 bp
*    41055    41765: contig of 711 bp in length
*    41766    41865: gap of 100 bp
*    41866    42599: contig of 734 bp in length
*    42600    42699: gap of 100 bp
*    42700    43424: contig of 725 bp in length
*    43425    43524: gap of 100 bp
*    43525    44232: contig of 708 bp in length
*    44233    44332: gap of 100 bp
*    44333    45042: contig of 710 bp in length
*    45043    45142: gap of 100 bp
*    45143    45846: contig of 704 bp in length
*    45847    45946: gap of 100 bp
*    45947    46654: contig of 708 bp in length
*    46655    46754: gap of 100 bp
*    46755    47467: contig of 713 bp in length
*    47468    47567: gap of 100 bp
*    47568    48283: contig of 716 bp in length
*    48284    48383: gap of 100 bp
*    48384    49091: contig of 708 bp in length
*    49092    49191: gap of 100 bp
*    49192    49895: contig of 704 bp in length
*    49896    49995: gap of 100 bp
*    49996    50672: contig of 677 bp in length
*    50673    50772: gap of 100 bp
*    50773    51486: contig of 714 bp in length
*    51487    51586: gap of 100 bp
*    51587    52304: contig of 718 bp in length
*    52305    52404: gap of 100 bp
*    52405    53131: contig of 727 bp in length
*    53132    53231: gap of 100 bp
*    53232    53919: contig of 688 bp in length
*    53920    54019: gap of 100 bp
*    54020    54727: contig of 708 bp in length
*    54728    54827: gap of 100 bp
*    54828    55545: contig of 718 bp in length

Alignment Scores:
Pred. No.:                1.31e+04        Length:         65263
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       14              Gaps:           0

US-09-858-332G-28 (1-6) x AC113185 (1-65263)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       7528 CACAACCACAACCACAAC 7545

RESULT 471
BX927211/c
LOCUS       BX927211               65572 bp    DNA     linear   VRT 23-FEB-2005
DEFINITION  Zebrafish DNA sequence from clone CH211-112A3, complete sequence.
ACCESSION   BX927211
VERSION     BX927211.4  GI:60219256
KEYWORDS    HTG.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 65572)
  AUTHORS   Holt,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 23, 2005 this sequence version replaced gi:58331573.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived
            Zebrafish pUC subclones occasionally display inconsistency over the
            length of mononucleotide A/T runs and conserved TA repeats. Where
            this is found the longest good quality representation will be
            submitted.
            Repeat names beginning 'Dr' were identified by the Recon repeat
            discovery system (Zhirong Bao and Sean Eddy, submitted), and those
            beginning 'drr' were identified by Rick Waterman (Stephen Johnson
            lab, WashU).  For further information see
            http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml CH211-112A3
            is from a CHORI-211 BAC library
            VECTOR: pTARBAC2.1.
FEATURES             Location/Qualifiers
     source          1. .65572
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-112A3"
                     /clone_lib="CHORI-211"
ORIGIN

Alignment Scores:
Pred. No.:              1.31e+04        Length:         65572
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     5               Gaps:           0

US-09-858-332G-28 (1-6) x BX927211 (1-65572)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      26987 CATAATCACAATCACAAT 26970

RESULT 472
AP007440
LOCUS       AP007440               66075 bp    DNA     linear   HTG 28-DEC-2004
DEFINITION  Lotus corniculatus var. japonicus clone LjT01F24, *** SEQUENCING IN
            PROGRESS ***, 15 unordered pieces.
ACCESSION   AP007440
VERSION     AP007440.1  GI:56805754
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1
  AUTHORS   Kaneko,T., Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Structural Analysis of a Lotus japonicus Genome. XI. Sequence
            Features and Mapping of Nine hundred twenty-one TAC Clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 66075)
  AUTHORS   Sato,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-2004) Shusei Sato, Kazusa DNA Research Institute,
            Department of Plant Gene Research; 2-6-7 Kazusa-kamatari, Kisarazu,
            Chiba, 292-0818, Japan (E-mail:ssato@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/, Tel:81-438-52-3935(ex.2337),
            Fax:81-438-52-3934)
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 15 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     7406: contig of 7406 bp in length
            *     7407     7506: gap of unknown length
            *     7507     8365: contig of 859 bp in length
            *     8366     8465: gap of unknown length
            *     8466     9112: contig of 647 bp in length
            *     9113     9212: gap of unknown length
            *     9213    10503: contig of 1291 bp in length
            *    10504    10603: gap of unknown length
            *    10604    11783: contig of 1180 bp in length
            *    11784    11883: gap of unknown length
            *    11884    13488: contig of 1605 bp in length
            *    13489    13588: gap of unknown length
            *    13589    14405: contig of 817 bp in length
            *    14406    14505: gap of unknown length
            *    14506    15909: contig of 1404 bp in length
            *    15910    16009: gap of unknown length
            *    16010    17608: contig of 1599 bp in length
            *    17609    17708: gap of unknown length
            *    17709    22148: contig of 4440 bp in length
            *    22149    22248: gap of unknown length
            *    22249    26647: contig of 4399 bp in length
            *    26648    26747: gap of unknown length
            *    26748    29919: contig of 3172 bp in length
            *    29920    30019: gap of unknown length
            *    30020    39280: contig of 9261 bp in length
            *    39281    39380: gap of unknown length
            *    39381    54135: contig of 14755 bp in length
            *    54136    54235: gap of unknown length
            *    54236    66075: contig of 11840 bp in length.
FEATURES             Location/Qualifiers
     source          1. .66075
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="genomic DNA"
                     /variety="japonicus"
                     /db_xref="taxon:34305"
                     /clone="LjT01F24"
                     /clone_lib="LjT library"
                     /note="TAC clone:TM0573, synonym:Lotus japonicus"
     gap             7407. .7506
                     /estimated_length=unknown
     gap             8366. .8465
                     /estimated_length=unknown
```

```
     gap             9113. .9212
                     /estimated_length=unknown
     gap             10504. .10603
                     /estimated_length=unknown
     gap             11784. .11883
                     /estimated_length=unknown
     gap             13489. .13588
                     /estimated_length=unknown
     gap             14406. .14505
                     /estimated_length=unknown
     gap             15910. .16009
                     /estimated_length=unknown
     gap             17609. .17708
                     /estimated_length=unknown
     gap             22149. .22248
                     /estimated_length=unknown
     gap             26648. .26747
                     /estimated_length=unknown
     gap             29920. .30019
                     /estimated_length=unknown
     gap             39281. .39380
                     /estimated_length=unknown
     gap             54136. .54235
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              1.32e+04        Length:         66075
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AP007440 (1-66075)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      40276 CATAATCACAATCATAAT 40293

RESULT 473
AC009199/c
LOCUS       AC009199               66498 bp    DNA     linear   HTG 20-SEP-1999
DEFINITION  Drosophila melanogaster strain y; cn bw sp chromosome 2 clone
            BACR10J23 (D1024) RPCI-98 10.J.23 map 37B-37B, *** SEQUENCING IN
            PROGRESS ***, 79 unordered pieces.
ACCESSION   AC009199
VERSION     AC009199.3  GI:5902283
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 66498)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 66498)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Sep 16, 1999 this sequence version replaced gi:5734692.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
            the following cutoffs: length >= 200 bases.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 79 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      571: contig of 571 bp in length
            *      572      651: gap of unknown length
            *      652     1378: contig of 727 bp in length
            *     1379     1458: gap of unknown length
            *     1459     2030: contig of 572 bp in length
            *     2031     2110: gap of unknown length
            *     2111     2716: contig of 606 bp in length
            *     2717     2796: gap of unknown length
            *     2797     3187: contig of 391 bp in length
            *     3188     3267: gap of unknown length
            *     3268     3883: contig of 616 bp in length
            *     3884     3963: gap of unknown length
            *     3964     4557: contig of 594 bp in length
            *     4558     4637: gap of unknown length
            *     4638     5327: contig of 690 bp in length
            *     5328     5407: gap of unknown length
            *     5408     5914: contig of 507 bp in length
            *     5915     5994: gap of unknown length
            *     5995     6552: contig of 558 bp in length
            *     6553     6632: gap of unknown length
            *     6633     7295: contig of 663 bp in length
            *     7296     7375: gap of unknown length
            *     7376     7993: contig of 618 bp in length
            *     7994     8073: gap of unknown length
            *     8074     8991: contig of 918 bp in length
            *     8992     9071: gap of unknown length
            *     9072     9658: contig of 587 bp in length
            *     9659     9738: gap of unknown length
            *     9739    10406: contig of 668 bp in length
            *    10407    10486: gap of unknown length
            *    10487    11164: contig of 678 bp in length
            *    11165    11244: gap of unknown length
            *    11245    11867: contig of 623 bp in length
            *    11868    11947: gap of unknown length
            *    11948    12561: contig of 614 bp in length
            *    12562    12641: gap of unknown length
            *    12642    13300: contig of 659 bp in length
            *    13301    13380: gap of unknown length
            *    13381    13979: contig of 599 bp in length
            *    13980    14059: gap of unknown length
            *    14060    14673: contig of 614 bp in length
            *    14674    14753: gap of unknown length
            *    14754    15366: contig of 613 bp in length
            *    15367    15446: gap of unknown length
            *    15447    16087: contig of 641 bp in length
            *    16088    16167: gap of unknown length
            *    16168    17458: contig of 1291 bp in length
            *    17459    17538: gap of unknown length
            *    17539    18499: contig of 961 bp in length
            *    18500    18579: gap of unknown length
            *    18580    19156: contig of 577 bp in length
            *    19157    19236: gap of unknown length
            *    19237    20512: contig of 1276 bp in length
```

```
*    20513    20592: gap of unknown length
*    20593    21219: contig of 627 bp in length
*    21220    21299: gap of unknown length
*    21300    21901: contig of 602 bp in length
*    21902    21981: gap of unknown length
*    21982    22674: contig of 693 bp in length
*    22675    22754: gap of unknown length
*    22755    23787: contig of 1033 bp in length
*    23788    23867: gap of unknown length
*    23868    25178: contig of 1311 bp in length
*    25179    25258: gap of unknown length
*    25259    26228: contig of 970 bp in length
*    26229    26308: gap of unknown length
*    26309    27229: contig of 921 bp in length
*    27230    27309: gap of unknown length
*    27310    28283: contig of 974 bp in length
*    28284    28363: gap of unknown length
*    28364    29769: contig of 1406 bp in length
*    29770    29849: gap of unknown length
*    29850    30989: contig of 1140 bp in length
*    30990    31069: gap of unknown length
*    31070    32741: contig of 1672 bp in length
*    32742    32821: gap of unknown length
*    32822    33963: contig of 1142 bp in length
*    33964    34043: gap of unknown length
*    34044    35604: contig of 1561 bp in length
*    35605    35684: gap of unknown length
*    35685    36676: contig of 992 bp in length
*    36677    36756: gap of unknown length
*    36757    37468: contig of 712 bp in length
*    37469    37548: gap of unknown length
*    37549    38942: contig of 1394 bp in length
*    38943    39022: gap of unknown length
*    39023    40372: contig of 1350 bp in length
*    40373    40452: gap of unknown length
*    40453    41945: contig of 1493 bp in length
*    41946    42025: gap of unknown length
*    42026    44246: contig of 2221 bp in length
*    44247    44326: gap of unknown length
*    44327    44957: contig of 631 bp in length
*    44958    45037: gap of unknown length
*    45038    45623: contig of 586 bp in length
*    45624    45703: gap of unknown length
*    45704    46099: contig of 396 bp in length
*    46100    46179: gap of unknown length
*    46180    46791: contig of 612 bp in length
*    46792    46871: gap of unknown length
*    46872    47504: contig of 633 bp in length
*    47505    47584: gap of unknown length
*    47585    48234: contig of 650 bp in length
*    48235    48314: gap of unknown length
*    48315    48934: contig of 620 bp in length
*    48935    49014: gap of unknown length
*    49015    49523: contig of 509 bp in length
*    49524    49603: gap of unknown length
*    49604    50176: contig of 573 bp in length
*    50177    50256: gap of unknown length
*    50257    50853: contig of 597 bp in length
*    50854    50933: gap of unknown length
*    50934    51493: contig of 560 bp in length
*    51494    51573: gap of unknown length
*    51574    52072: contig of 499 bp in length
*    52073    52152: gap of unknown length
*    52153    52746: contig of 594 bp in length
*    52747    52826: gap of unknown length
*    52827    53416: contig of 590 bp in length
*    53417    53496: gap of unknown length
*    53497    53966: contig of 470 bp in length
*    53967    54046: gap of unknown length
*    54047    54656: contig of 610 bp in length
*    54657    54736: gap of unknown length
*    54737    55365: contig of 629 bp in length
*    55366    55445: gap of unknown length
*    55446    56028: contig of 583 bp in length
*    56029    56108: gap of unknown length
*    56109    56716: contig of 608 bp in length
*    56717    56796: gap of unknown length
*    56797    57425: contig of 629 bp in length
*    57426    57505: gap of unknown length
*    57506    58065: contig of 560 bp in length
*    58066    58145: gap of unknown length
*    58146    58782: contig of 637 bp in length
*    58783    58862: gap of unknown length
*    58863    59502: contig of 640 bp in length
*    59503    59582: gap of unknown length
*    59583    60193: contig of 611 bp in length
*    60194    60273: gap of unknown length
*    60274    60890: contig of 617 bp in length
*    60891    60970: gap of unknown length
*    60971    61575: contig of 605 bp in length
*    61576    61655: gap of unknown length
*    61656    62271: contig of 616 bp in length
*    62272    62351: gap of unknown length
*    62352    62982: contig of 631 bp in length
*    62983    63062: gap of unknown length
*    63063    63693: contig of 631 bp in length

Alignment Scores:
Pred. No.:                1.33e+04        Length:          66498
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       14              Gaps:            0

US-09-858-332G-28 (1-6) x AC009199 (1-66498)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      64963 CACAATCACAACCACAAT 64946

RESULT 474
AC069130/c
LOCUS       AC069130               66874 bp    DNA     linear   PRI 09-MAY-2001
DEFINITION  Homo sapiens BAC clone RP11-468D10 from Y, complete sequence.
ACCESSION   AC069130
VERSION     AC069130.6  GI:13435287
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 66874)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
   PUBMED   9847074
REFERENCE   2  (bases 1 to 66874)
  AUTHORS   Yoakum,M., Maupin,R. and Kisic,M.
  TITLE     The sequence of Homo sapiens BAC clone RP11-468D10
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 66874)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAY-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 66874)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAR-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 66874)
  AUTHORS   Waterston,R.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Mar 23, 2001 this sequence version replaced gi:10312294.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0468D10
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            The position of this clone was established as part of a
            collaboration between the Human Chromosome Y Mapping Project
            (Tomoko Kawaguchi, Helen Skaletsky, Laura G. Brown, Steve Rozen,
            and David C. Page at the Whitehead Institute for Biomedical
            Research, Cambridge MA) and the Washington University Genome
            Sequencing Center, St. Louis MO.

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries.  Genomics 51:1-8.  The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu)
            VECTOR:  pBACe3.6
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-157F24, 200 bp overlap; the
            clone sequenced to the right is RP11-945E12, 200 bp overlap.
            Actual start of this clone is at base position 41625 of
            RP11-157F24; actual end is at base position 66680 of RP11-468D10.
FEATURES             Location/Qualifiers
     source          1. .66874
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="Y"
                     /map="Y"
                     /clone="RP11-468D10"
                     /clone_lib="RPCI-11"
     repeat_region   1220. .1438
                     /rpt_family="MaLR"
     repeat_region   1767. .1989
                     /rpt_family="MaLR"
     repeat_region   3126. .3379
                     /rpt_family="Alu"
     repeat_region   3987. .4296
                     /rpt_family="Alu"
     repeat_region   5063. .5373
                     /rpt_family="Alu"
     repeat_region   5681. .5939
                     /rpt_family="L1"
     repeat_region   5940. .6253
                     /rpt_family="Alu"
     misc_feature    6133. .6255
                     /note="similar to  EST AI866821 (NID:g5530852) tz53f11.x1"
     repeat_region   6254. .8371
                     /rpt_family="L1"
     repeat_region   8384. .8644
                     /rpt_family="L1"
     repeat_region   8648. .10180
                     /rpt_family="L1"
     repeat_region   11129. .11422
                     /rpt_family="Alu"
     repeat_region   11693. .11978
                     /rpt_family="Alu"
     repeat_region   13327. .14033
                     /rpt_family="L1"
     repeat_region   14070. .14451
                     /rpt_family="L1"
     misc_feature    14357. .14664
                     /note="similar to  EST AI866821 (NID:g5530852) tz53f11.x1"
     repeat_region   14498. .14860
                     /rpt_family="L1"
     misc_feature    14840. .15020
                     /note="similar to  EST AI866821 (NID:g5530852) tz53f11.x1"
     repeat_region   15105. .16254
                     /rpt_family="L1"
     repeat_region   16257. .16375
                     /rpt_family="L1"
     repeat_region   16393. .16565
                     /rpt_family="Alu"
     repeat_region   16574. .17833
                     /rpt_family="L1"
     repeat_region   17839. .18268
                     /rpt_family="ERV1"
     repeat_region   18297. .18361
                     /rpt_family="ERV1"
     repeat_region   25846. .26717
                     /rpt_family="ERV1"
     repeat_region   26814. .27093
                     /rpt_family="ERVK"
     repeat_region   27102. .28327
                     /rpt_family="ERVK"
     repeat_region   28331. .28766
                     /rpt_family="ERVK"
     repeat_region   28793. .28983
                     /rpt_family="Alu"
     repeat_region   29096. .29184
                     /rpt_family="L1"
     repeat_region   29537. .29847
                     /rpt_family="Alu"
     repeat_region   30646. .31234
                     /rpt_family="L1"
     repeat_region   31464. .31574
                     /rpt_family="L1"
     repeat_region   31569. .31689
                     /rpt_family="L1"
     repeat_region   31700. .31874
                     /rpt_family="L1"
     repeat_region   31883. .32100
                     /rpt_family="ERV1"
     repeat_region   32241. .32962
                     /rpt_family="L1"
     repeat_region   37763. .37862
                     /rpt_family="L1"
     repeat_region   37863. .38156
                     /rpt_family="Alu"
     repeat_region   38157. .38224
                     /rpt_family="L1"
     misc_feature    38354. .38731
                     /note="similar to  EST W63784 (NID:g1371385) zd30f11.s1"
     misc_feature    38359. .38731
                     /note="similar to  EST AW959166 (NID:g8148850)"
     misc_feature    38362. .38731
                     /note="similar to  EST AU035808 (NID:g3718816)"
     misc_feature    38474. .38731
                     /note="similar to  EST AI248008 (NID:g3843405) qh63d04.x1"
```

```
     misc_feature    38491. .38731
                     /note="similar to  EST AA460930 (NID:g2186050) zx61c04.s1"
     misc_feature    38551. .38731
                     /note="similar to  EST AA100084 (NID:g1647473) zk60a09.r1"
     misc_feature    38771. .39051
                     /note="similar to  EST AU035808 (NID:g3718816)"
     misc_feature    38771. .39023
                     /note="similar to  EST AW959166 (NID:g8148850)"
     misc_feature    38771. .39011
                     /note="similar to  EST AA100084 (NID:g1647473) zk60a09.r1"
     misc_feature    38771. .39011
                     /note="similar to  EST AI248008 (NID:g3843405) qh63d04.x1"
     misc_feature    38771. .39010
                     /note="similar to  EST W63784 (NID:g1371385) zd30f11.s1"
     misc_feature    38771. .39007
                     /note="similar to  EST AA460930 (NID:g2186050) zx61c04.s1"
     repeat_region   39286. .39550
                     /rpt_family="Alu"
     repeat_region   39929. .40106
                     /rpt_family="Alu"
     repeat_region   40116. .40555

Alignment Scores:
Pred. No.:              1.34e+04        Length:         66874
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     8               Gaps:           0

US-09-858-332G-28 (1-6) x AC069130 (1-66874)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      11576 CATAATCATAATCATAAT 11559

RESULT 475
AC017241
LOCUS       AC017241               67847 bp    DNA     linear   HTG 09-DEC-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC017241
VERSION     AC017241.1  GI:6553745
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 67847)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10209995 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .67847
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
ORIGIN

Alignment Scores:
Pred. No.:              1.36e+04        Length:         67847
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AC017241 (1-67847)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      32474 CACAATCACAACCACAAC 32491

RESULT 476
AP003399/c
LOCUS       AP003399               68801 bp    DNA     linear   PRI 30-MAR-2002
DEFINITION  Homo sapiens genomic DNA, chromosome 11q clone:RP11-448N3, complete
            sequence.
ACCESSION   AP003399
VERSION     AP003399.2  GI:19879821
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA
  JOURNAL   Published Only in Database (2001)
REFERENCE   2  (bases 1 to 68801)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-2001) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     On Apr 1, 2002 this sequence version replaced gi:13366103.
FEATURES             Location/Qualifiers
     source          1. .68801
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-448N3"
ORIGIN

Alignment Scores:
Pred. No.:              1.38e+04        Length:         68801
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     8               Gaps:           0

US-09-858-332G-28 (1-6) x AP003399 (1-68801)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       3680 CATAATCACAACCATAAT 3663

RESULT 477
AC087771
LOCUS       AC087771               69590 bp    DNA     linear   PLN 13-JUL-2001
DEFINITION  Genomic Sequence for Medicago truncatula, clone 8D15, complete
            sequence.
ACCESSION   AC087771
VERSION     AC087771.3  GI:14717388
KEYWORDS    HTG.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
```

```
REFERENCE   1  (bases 1 to 69590)
  AUTHORS   Huang,E., de la Bastide,M., Nascimento,L., Spiegel,L., Kirchoff,K.,
            Preston,R., King,L., Zutavern,T., Santos,L., Miller,B., Kuit,K.,
            Cunnius,D.M., Bell,M., Balija,V., Shah,R., Bahret,A., Yang,C.,
            Palmer,L., O'Shaughnessy,A., Dedhia,N. and McCombie,W.R.
  TITLE     Genomic Sequence for Medicago truncatula, clone 8D15, complete
            sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 69590)
  AUTHORS   McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JAN-2001) Lita Annenberg Hazen Genome Sequencing
            Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
            Harbor, NY 11724, USA
REFERENCE   3  (bases 1 to 69590)
  AUTHORS   McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUN-2001) Lita Annenberg Hazen Genome Sequencing
            Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
            Harbor, NY 11724, USA
REFERENCE   4  (bases 1 to 69590)
  AUTHORS   Huang,E., de la Bastide,M., Spiegel,L., Nascimento,L., Kirchoff,K.,
            Preston,R., King,L., Zutavern,T., Santos,L., Miller,B., Kuit,K.,
            Cunnius,D.M., Bell,M., Balija,V., Shah,R., Bahret,A., Yang,C.,
            Dike,S., O'Shaughnessy,A., Palmer,L., Dedhia,N. and McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2001) Lita Annenberg Hazen Genome Center, Cold
            Spring Harbor Laboratory, 1 Bungtown Rd, Cold Spring Harbor, NY
            11724, USA
COMMENT     On Jul 13, 2001 this sequence version replaced gi:14578148.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest.
FEATURES             Location/Qualifiers
     source          1. .69590
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3880"
                     /clone="8D15"
     misc_feature    47170. .47550
                     /note="The sequence is covered by multiple subclones
                     derived from a PCR product amplified with a high fidelity
                     polymerase.  The sequence is of high quality."
     misc_feature    47550. .48613
                     /note="The sequence is covered by a PCR product which was
                     amplified with a high fidelity polymerase and by 2 M13
                     subclones.  The sequence is high quality."
     misc_feature    48613. .48750
                     /note="The region is covered only by PCR products which
                     were amplified with a high fidelity polymerase and
                     sequenced with both dye terminator and GTP terminator
                     chemistries."
     misc_feature    48750. .49160
                     /note="The sequence is covered by a PCR product which was
                     amplified with a high fidelity polymerase and one M13
                     subclone.  The sequence is high quality."
     misc_feature    49160. .49270
                     /note="The region is covered only by a PCR product which
                     was amplified with a high fidelity polymerase,as well as
                     sequences from a subcloned PCR product.  The sequence is
                     high quality."
     misc_feature    50440. .50630
                     /note="The sequence is covered only by a PCR product which
                     was amplified by a high fidelity polymerase."
     misc_feature    50655. .50665
                     /note="The sequence is covered by a single high quality
                     M13 subclone which was sequenced with both dye primer and
                     dye terminator chemistries."
ORIGIN
```

```
Alignment Scores:
Pred. No.:              1.39e+04        Length:         69590
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x AC087771 (1-69590)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      64566 CACAATCATAATCACAAC 64583

RESULT 478
AL929145
LOCUS       AL929145               70534 bp    DNA     linear   ROD 10-APR-2003
DEFINITION  Mouse DNA sequence from clone RP23-362L7 on chromosome 2, complete
            sequence.
ACCESSION   AL929145
VERSION     AL929145.12  GI:29786549
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 70534)
  AUTHORS   Pelan,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-APR-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 10, 2003 this sequence version replaced gi:29691849.
            Sequence from the Mouse Genome Sequencing Consortium whole genome
            shotgun may have been used to confirm this sequence.  Sequence data
            from the whole genome shotgun alone has only been used where it has
            a phred quality of at least 30.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-362L7 is
            from the RPCI-23 Mouse BAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .70534
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
```

```
                     /chromosome="2"
                     /clone="RP23-362L7"
                     /clone_lib="RPCI-23"
ORIGIN

Alignment Scores:
Pred. No.:                1.41e+04        Length:          70534
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       9               Gaps:            0

US-09-858-332G-28 (1-6) x AL929145 (1-70534)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      32873 CATAATCATAATCATAAT 32890

RESULT 479
AB024032/c
LOCUS       AB024032               70670 bp    DNA     linear   PLN 14-FEB-2004
DEFINITION  Arabidopsis thaliana genomic DNA, chromosome 5, TAC clone:K9P8.
ACCESSION   AB024032 BA000015
VERSION     AB024032.1  GI:4519191
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E., Kotani,H.
            and Tabata,S.
  TITLE     Structural analysis of Arabidopsis thaliana chromosome 5. X.
            Sequence features of the regions of 3,076,755 bp covered by sixty
            P1 and TAC clones
  JOURNAL   DNA Res. 7 (1), 31-63 (2000)
   PUBMED   10718197
REFERENCE   2  (bases 1 to 70670)
  AUTHORS   Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-FEB-1999) Yasukazu Nakamura, Kazusa DNA Research
            Institute, Department of Plant Gene Research; 1532-3, Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
            Tel:81-438-52-3935, Fax:81-438-52-3934)
COMMENT     Address for correspondence: kaos@kazusa.or.jp
            For the latest information on annotation of this clone, please see
            http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=K9P8
            Genes with similarity to proteins in the databases are described in
            'product' or 'note' qualifiers.  Genes that have no significant
            protein similarity are described as 'unknown protein'.
            The software programs used to predict genes include: Grail
            (Informatics Group, Oak Ridge National Laboratory,
            http://compbio.ornl.gov/Grail-1.3/),
            GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
            NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
            Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
            SplicePredictor (Volker Brendel, Stanford University,
            http://gremlin1.zool.iastate.edu/cgi-bin/sp.cgi).
            Genes encoding tRNAs are predicted by tRNAscan-SE
            (Sean Eddy, Washington University School of Medicine, St. Louis,
            http://genome.wustl.edu/eddy/tRNAscan-SE/).
            This sequence may not be the entire insert of this clone. It may be
            shorter because we remove overlaps between neighboring submissions.
            The 5' clone is K21G20 and the 3' clone is MPF21.
FEATURES             Location/Qualifiers
     source          1. .70670
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /chromosome="5"
                     /clone="K9P8"
                     /clone_lib="Mitsui TAC"
                     /ecotype="Columbia"
     CDS             complement(join(307. .564,665. .1123,1199. .1639,1746. .2189,
                     2305. .2508))
                     /note="gene_id:K9P8.1"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="myrosinase binding protein-like"
                     /protein_id="BAA97008.1"
                     /db_xref="GI:8777418"
                     /translation="MTERLEAEGSRRGNNSKWDDKSDHDDVTKIYVNYSLSGIESIRF
                     DYVKSGKPKDGPFRGQSYNTYTHTFEINHLKYEHLESVEGYYTEDTGIEALQFKTNLR
                     ISEPIGYHHDGCTKFILAVEGKKIIGFHGSSFALRVGSLGAYFTWDTPTIIKAVGGKV
                     GTKWDDGVNQAGFTKIHVRSGQKGIQFIKLEYVDKDGNLTDGPIHGSIYRRGSPHVFE
                     FKHDEEYLVSVEGYYEGDEECEVIQGLQFRTNIKTSELMGSNTGKKFKLTASGMKIVG
                     FHGYAEKNLSSLGAYLTPLTPKKSECKGITDGSNVWNWDDGAFEGIRKVSVFFTALRI
                     RCLMINYEDAGKVVKRSHGLNNNDHQEEKFVVDYPNEFITSVVGTMSSSSVMPLIFKT
                     SKGRTSKQFGDYSFDDSVEFVLESKGCAIVGFHGWHNPVSGYMTALGAYYYPMPLPPA
                     AEKLEAQGGAGGAPWDDGSNFQGVRKIYIGTGEVGIVSIKFLYENDVHEIVVGDHHGN
                     KNLLRHEEFDLDYPSEYLTSVEGSYDVVPGSEEYEVIIMLKFTTNKRTSPCYGLDDDP
                     IFVLHKEGHKIVGFHGKSSNMLHKLGIHVLPISNS"
     CDS             join(12265. .12459,12536. .12635,12736. .12896,12981. .13178,
                     13276. .13368,13474. .13590,14173. .14420,14770. .14941,
                     15638. .15748,15833. .16036,16291. .16453,16741. .16831,
                     16928. .16997,17311. .17439,17523. .17561,17646. .17735)
                     /note="gene_id:K9P8.2"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="mitotic checkpoint protein-like"
                     /protein_id="BAA97009.1"
                     /db_xref="GI:8777419"
                     /translation="MILRTPQPKRLRSDAGESPFPTGATGSGNQLIIYEDSPLPAPAP
                     LQTSHDHSADQHLCTYQCRQMVKADVLDALSTAEKQVEESKTKLQTLNAANFTEADAER
                     KHFRDKFLYSEQELAAAKAREKMLQEQLLMEINNSQERYTKELQSCHELEVKLQNEMN
                     LRKKAESSAATAEEKAKLLEDKLTQLSGSVDREKKRLNNDIAQLGKEAKLSVARIGAD
                     LERMQCRAQNAETESNLLRSQLEHLKLIFDECLQEKTEVDKKLSSFTSEAASSSDNSV
                     LVKHLQEELKRYEAEVREARKLKSRHLDAELLNVNLLEEQSRRERAESELSKFHDLQL
                     SMEKLENELSSWKSLLNDIPGVSCPDDIVMRFSVLQNEVVQSTMKIGEASTRIKQLEE
                     TLEAIQLGRQNAVSEAALAKEKSEALKTDVKRIEVMLTLVTEEKEQLKAVVNELRKSN
                     SEGSVSGAADGALIQGFESSLAKKENYIKDLEQDLNQLKDVNNRQRTEIELLNEKLVD
                     EARRNKSLERDSDRLRSEISLLESKLGHGDYSAANTRVLRMVNTLGVENEAKQTIEAL
                     QAELQKTKERLQAVEELKSQSGDAGKLVDSHITGKIAQLKEQNATLEKREERYKTVFA
                     DRISVFRRACCELFGYKIVMDEHQRPNGIPVTRFTLQSIYAQSDDEKLEFEYESGNTS
                     ILNNEYASQGDIAKQIEIFIRKFNSIPAFTANLTMESFNRRTLY"
     CDS             complement(join(18449. .18538,18615. .19746,19869. .20138,
                     20450. .20712,20824. .21001,21198. .21398,21898. .22103))
                     /note="emb|CAA70310.1
                     gene_id:K9P8.3"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="chloride channel"
                     /protein_id="BAA97010.1"
                     /db_xref="GI:8777420"
                     /translation="MDDRHEGDHHDIEVEGGALHGFERKISGILDDGSVGFRQPLLAR
                     NRKNTTSQIAIVGANTCPIESLDYEIFENDFFKQDWRSRKKIEILQYTFLKWALAFLI
                     GLATGLVGFLNNLGVENIAGFKLLLIGNLMLKEKYFQAFFAFAGCNLILATAAASLCA
                     FIAPAAAGSGIPEVKAYLNGIDAYSILAPSTLFVKIFGSIFGVAAGFVVGKEGPMVHT
                     GACIANLLGQGGSKKYRLTWKWLRFFKNDRDRRDLITCGAAAGVAAAFRAPVGGVLFA
                     LEEAASWWRNALLWRTFFTTAVVAVVLRSLIEFCRSGRCGLFGKGGLIMFDVNSGPVL
                     YSTPDLLAIVFLGVIGGVLGSLYNYLVDKVLRTYSIINEKGPRFKIMLVMAVSILSSC
                     CAFGLPWLSQCTPCPIGIEEGKCPSVGRSSIYKSFQCPPNHYNDLSSLLLNTNDDAIR
                     NLFTSRSENEFHISTLAIFFVAVYCLGIITYGIAIPSGLFIPVILAGASYGRLVGRLL
                     GPVSQLDVGLFSLLGAASFLGGTMRMTVSLCVILLELTNNLLMLPLVMLVLLISKTVA
                     DCFNRGVYDQIVTMKGLPYMEDHAEPYMRNLVAKDVVSGALISFSRVEKVGVIWQALK
                     MTRHNGFPVIDEPPFTEASELCGIALRSHLLVLLQGKKFSKQRTTFGSQILRSCKARD
                     FGKAGLGKGLKIEDLDLSEEEMEMYVDLHPITNTSPYTVLETLSLAKAAILFRQLGLR
                     HLCVVPKTPGRPPIVGILTRHDFMPEHVLGLYPHIDPLK"
     CDS             complement(join(27195. .27492,27560. .27850,27936. .28866,
                     28957. .29018,29102. .29174,29259. .29353,29396. .29577,
                     29672. .29816,30379. .30482,30541. .30676,30763. .30833,
                     30955. .31054,31232. .31302,31401. .31457,31569. .31673,
                     31914. .31979))
                     /note="unnamed protein product; gene_id:K9P8.4
```

```
                     pir||T04010
                     strong similarity to unknown protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAA97011.1"
                     /db_xref="GI:8777421"
                     /translation="MFEEKIMDIGEDVKPLNSSDTKVDPAVPASLTWQRKIDSDVKAP
                     REFNLSVKEIFQLAPVGIRLWFLCREEAAKGRLAFIDPFSKHSVTSSHGVPLGGIGAG
                     SIGRSFKGEFQRWQLFPPKCEDEPVLANQFSAFVSRANGKKYSSVLCPRNPKLDKQDS
                     ESGIGSWDWNLKGDKSTYHALYPRSWTMYEGNIYAFEYSFTLVNLIQSLGLPVVKFHP
                     LYLTTTKKAAFQCLFSLSQTANGLPSLSYAISAQATDGVSVSACPFFIVSGKQDGITA
                     KDMWQAVKENGSFDHLKASEASMQSDHGSSIGAAVAASVTVLPGESRIVTFSLAWDCP
                     EVQFPSGKIYSRQPIKQNLFRRYTKFYGNNGDAAAQIAHDAILGHSQWESWIEDWQRP
                     ILEDKRLPAWYPVTLFNELYYLNSGGTLWTDGSSPVHSLAGVREKKFSLDKSQLGLKN
                     DIDVPHQNDTAVSVLEKMASTLEELHASTTSNSAFGTKLLEEGEENIGHFLYLEGIEY
                     RMWNTYDVHFYASFALVMLFPKLELSIQRDFAAAVMLHDPTKVKTLSEGQWVQRKVLG
                     AVPHDLGINDPWFEVNGYTLHNTDRWKDLNPKFVLQVYRDVVATGDKKFASAVWPSVY
                     VAMAYMAQFDKDGDGMIENEGFPDQTYDTWSASGVSAYCGGLWVAALQAASALARVVG
                     DKNSQDYFWSKFQKAKVVYEKKLWNGSYFNYDNSGSQYSSTIQADQLAGQWYARASGL
                     LPIVDEDKARTALEKVYNYNVMKIKDGKRGAVNGMHPNGKVDTASMQSREIWSGVTYA
                     LSATMIQEGLVEMAFQTASGIYEAAWSETGLGYSFQTPESWNTVDEYRSLTYMRPLAI
                     WAMQWALTKTSQKQEQLGLEPEQQEPELEPSSSMKHDIGFSRVSRLLSLPNEASAKST
                     LQTLFDYTCRRMMS"
CDS                  join(33430. .34011,34089. .34341,34451. .34530,34614. .34756,
                     34837. .34935,35051. .35240,35350. .35748,35845. .36255)
                     /note="gb|AAF27639.1
                     gene_id:K9P8.5"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="heat shock protein 70"
                     /protein_id="BAA97012.1"
                     /db_xref="GI:8777422"
                     /translation="MASSAAQIHILGGIGFPTSSSSSSTKNLDNKTNSIPRSVFFGNR
                     TSPFTTPTSAFLRMGRRNNNASRYTVGPVRVVNEKVVGIDLGTTNSAVAAMEGGKPTI
                     VTNAEGQRTTPSVVAYTKSKDRLVGQIAKRQAVVNPENTFFSVKRFIGRRMNEVAEES
                     KQVSYRVIKDENGNVKLDCPAIGKQFAAEEISAQVLRKLVDDASRFLNDKVTKAVITV
                     PAYFNDSQRTATKDAGRIAGLEVLRIINEPTAASLAYGFERKSNETILVFDLGGGTFD
                     VSVLEVGDGVFEVLSTSGDTHLGGDDFDKRVVDWLASTFKKDEGIDLLKDKQALQRLT
                     EAAEKAKIELSSLTQTNMSLPFITATADGPKHIETTLTRGKFEELCSDLLDRVRTPVE
                     NSLRDAKLSFKDIDEVILVGGSTRIPAVQDLVRKLTGKEPNVSVNPDEVVALGAAVQA
                     GVLSGDVSDIVLLDVTPLSLGLETLGGVMTKIIPRNTTLPTSKSEVFSTAADGQTSVE
                     INVLQGEREFVRDNKSIGSFRLDGIPPAPRGVPQIEVKFDIDANGILSVSASDKGTGK
                     KQDITITGASTLPKDEVDTMVQEAERFAKEDKEKRDAIDTKNQADSVVYQTEKQLKEL
                     GEKIPGPVKEKVEAKLQELKEKIASGSTQEIKDTMAALNQEVMQIGQSLYNQPQPGGA
                     DSPPGGEASSSSDTSSSAKGGDNGGDVIDADFTDSN"
CDS                  complement(join(36597. .37318,37399. .37543))
                     /note="unnamed protein product; gene_id:K9P8.6
                     unknown protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAA97013.1"
                     /db_xref="GI:8777423"
                     /translation="MAEERSPSKVKFMCSFGGRILPRPSDSVLKYVGGETRVVAVSPD
                     ISFSELMKKLTAITENDIVLKYQIIPEDLDALVSVKSDEDVKHMIEEYNRHETPKLRT
                     FLFPANPVVLESQLGPIEPQTIEQRYIEAINGILRTSKSATALRAPIKTRPSFTVSTC
                     SSPKSESPDGYSHEPPETSFQNNYQLSRLYPMHRVHSSPNISHQLQPHSHYHNHNHNA
                     YVQPPTNYLTCRPRLPTPLEIPRGIGWDQSYTTTHNSGGGNGKYGCNDDRRFWGRASS
                     VPQSPRNHGLRL"
CDS                  complement(join(38220. .38416,38497. .38790,38900. .38981,
                     39082. .39933,40026. .40131,40224. .40378,40465. .40617,
                     40704. .40997,41108. .41293,41386. .41869,41987. .42111,
                     42197. .42293,42379. .42491,42592. .42696))
                     /note="unnamed protein product; gb|AAF56406.1
                     gene_id:K9P8.7
                     strong similarity to unknown protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAA97014.1"
                     /db_xref="GI:8777424"
                     /translation="MVKVRMNTADVAAEVKCLKRLIGMRCSNVYDISPKTYMFKLLNS
                     SGITESGESEKVLLLMESGVRLHTTAYVRDKSNTPSGFTLKLRKHIRTRRLEDVRQLG
                     YDRIIVFQFGLGANAHYVILELYAQGNIILTDSEYMIMTLLRSHRDDNKGFAIMSRHR
                     YPIEICRVFERTTVSKLQESLTAFVLKDHDAKQIEPKEQNGGKKGGKSNDSTGAKQYT
                     LKNILGDALGYGPQLSEHIILDAGLVPTTKLSEDKKLDDNEIQLLVQAVIVFEDWLED
                     IINGQKVPEGYILMQKQILANDTTSESGGVKKMYDEFCSILLNQFKSRVYEKFETFDA

Alignment Scores:
Pred. No.:              1.41e+04        Length:         70670
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x AB024032 (1-70670)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      36830 CACAACCACAACCACAAT 36813

RESULT 480
AP001300/c
LOCUS       AP001300               71327 bp    DNA     linear   PLN 14-FEB-2004
DEFINITION  Arabidopsis thaliana genomic DNA, chromosome 3, BAC clone:F5N5.
ACCESSION   AP001300 BA000014
VERSION     AP001300.1  GI:7209734
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Kaneko,T., Katoh,T., Sato,S., Nakamura,A., Asamizu,E. and Tabata,S.
  TITLE     Structural analysis of Arabidopsis thaliana chromosome 3. II.
            Sequence features of the 4,251,695 bp regions covered by 90 P1, TAC
            and BAC clones
  JOURNAL   DNA Res. 7 (3), 217-221 (2000)
   PUBMED   10907853
REFERENCE   2  (bases 1 to 71327)
  AUTHORS   Kaneko,T., Kato,T., Sato,S., Nakamura,Y., Asamizu,E. and Tabata,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2000) Yasukazu Nakamura, Kazusa DNA Research
            Institute, Department of Plant Gene Research; 1532-3, Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
            Tel:81-438-52-3935, Fax:81-438-52-3934)
COMMENT     Address for correspondence: kaos@kazusa.or.jp
            For the latest information on annotation of this clone, please see
            http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=F5N5
            Genes with similarity to proteins in the databases are described in
            'product' or 'note' qualifiers.  Genes that have no significant
            protein similarity are described as 'unknown protein'.
            The software programs used to predict genes include: Grail
            (Informatics Group, Oak Ridge National Laboratory,
            http://compbio.ornl.gov/Grail-1.3/),
            GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
            NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
            Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
            SplicePredictor (Volker Brendel, Stanford University,
            http://gremlin1.zool.iastate.edu/cgi-bin/sp.cgi).
            Genes encoding tRNAs are predicted by tRNAscan-SE
            (Sean Eddy, Washington University School of Medicine, St. Louis,
            http://genome.wustl.edu/eddy/tRNAscan-SE/).
            This sequence may not be the entire insert of this clone. It may be
            shorter because we remove overlaps between neighboring submissions.
            The 5' clone is MWI23 and the 3' clone is MXC7.
FEATURES              Location/Qualifiers
     source           1. .71327
                      /organism="Arabidopsis thaliana"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:3702"
                      /chromosome="3"
                      /clone="F5N5"
                      /clone_lib="IGF BAC"
                      /ecotype="Columbia"
     CDS              join(1008. .1207,1484. .1724,1811. .1931,2032. .2216)
```

/note="unnamed protein product; contains similarity to unknown protein
gb|AAF35411.1
gene_id:F5N5.2"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB03030.1"
/db_xref="GI:11994714"
/translation="MLAIFDKNVAKTPEALQGQEGGSVCALKDRFLPNHFSSVYPGAV
TINLGSSGFIACSLEKQNPLLPRLFAVVDDMFCIFQGHIENVPILKQQYGLTKTATEV
TIVIEAYRTLRDRGPYSAEQVVRDFQGKFGFMLYDCSTQNVFLAGDVDGSVPLYWGTD
AEGHLVVSDDVETVKKGCGKSFAPFPKGCFFTSSGGLRSYEHPSNELKPVPRVDSSGE
VCGVTFKVDSEAKKEAMPRVGSVQNWSKQI"
CDS complement(join(2595. .3359,3443. .5080))
/note="gene_id:F5N5.3"
/codon_start=1
/evidence=not_experimental
/product="eukaryotic translation initiation factor 3 subunit 8-like protein"
/protein_id="BAB03031.1"
/db_xref="GI:11994715"
/translation="MFRLQSGSESEDESDFEKEEVEGNNTRYLADLPDGEICCGMDTK
RVVKPKKDKRFEEMANTIENMKHAMNINDCVYLQETFEKLNKQISKSVKTPTLYIKTL
VMLEDFLNEDNMKTKEKMSTSNSKALNAMRQKLKKNNLQYQEDIKRFRESPEIEDDDE
YEEEVVEDSADNVSWEMLFSLDHEEITWNMVNKKFKEIRAARWSKRRSSSLKLKPGET
HAQKHMDLTKIAKTPAQKVEILFSVISAEFNVNSGGLSGYMPIDVWKKCVVNMLTILD
ILVKYYNIVVDDTVEPDEKETSKPAAYDGTIRVSGNLVAFLEKIETEFFKSLQCIDPH
TNDYVERLKDEPMPLALAQSIQDYLERTGDSKAASKVAFILVESIYYKPQEVFDAMSK
LADEEIEEANEESGSSSSSFIVVAEIVPRKPTFAKSSRAMMDTLVSFIYKNGDERTKA
RAMLCDIYQHALMDNFVTARDLLLMSHLQENIQHMDISTQILFNRTMAQLGLCAFRVG
MITESHSCLSELYSGNRVRELLGQGVSQSRDHEKTTEQMLMERRTQIPYHMNINLELL
EAVYLTCAMLLEVPNMAANSHDSKHKPISKNIQRLLEKSEKQAFTAPPENVRVHVIAA
TRALIKGNFQEAFSVLNSLDIWRLFKNRDSILDMVKASISEVALRTYLFTYSSSCYKS
LSLAELAKMFDISESHVYSIVSKMMINKELNATWDQPTQCIVFHEVQHNRAQSLAFQI
TEKLATLAESNESAMESKTGGVGLDMTSRRRENSQDYAAAASDNRGLKMDGSSLTRGV
SA"
CDS 8171. .9412
/note="unnamed protein product; gb|AAC67359.1
gene_id:F5N5.4
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB03032.1"
/db_xref="GI:11994716"
/translation="MTLTISDLPRDLKKKIFSRIPLRYVRALRLTCKEWETLIKSRSL
KIDEEESQMVALMDYNLCLMSKSFNGGDPSTEIKGKLTCFDEQVKVSMLFHCEGLLLC
ILKDDNTKVVVWNPYLGQRRFIQVRFSHIPGGWDQFSYALGYYKNNNMMKSFKLLRFF
DYFYTSNDFVFYEIYDFDSGLWKTLDVTPYWRIRFFDDIGVSLKGNTYWCAKERKPEW
NADPFIIVDHIICFDFTSERFGPLVPLPFRAMKESLVTLSCVREEKLAALFCRDVVVE
VWITTTIEFDKVSWSKFLTFGVRDCPMIYSCTGSFFIDEVNKVAVIFGQPLKRDKVFI
IGEAGQVRDVDLGEPAHQDSWLYVCPYVPSLVKIKQPPGCKRRRQSSSENRRYKRNMM
RLVELEKYHIAMEPKRGDKTI"
CDS complement(join(9889. .9933,10033. .10199,10308. .10386,
10471. .10539,10626. .10718,10802. .10868,10970. .11031,
11191. .11236,11314. .11359,11443. .11500,11613. .11651,
11741. .11812,11906. .11997,12413. .12476))
/note="gene_id:F5N5.5"
/codon_start=1
/evidence=not_experimental
/product="AtDMC1 (meiotic recombination protein)-like protein"
/protein_id="BAB03033.1"
/db_xref="GI:11994717"
/translation="MQLVEREENDEDEDLFEMIDKLIAQGINAGDVKKLQEAGIHTCN
GLMMHTKKNLTGIKGLSEAKVDKICEAAEKIVNFGYMTGSDALIKRKSVVKITTGCQA
LDDLLGGGIETSAITEAFGEFRSGKTQLAHTLCVTTQLPTNMKGGNGKVAYIDTEGTF
RPDRIVPIAERFGMDPGAVLDNIIYARAYTYEHQYNLLLGLAAKMSEEPFRILIVDSI
IALFRVDFTGRGELADRQQKLAQMLSRLIKIAEEFNVAVYMTNQVIADPGGGMFISDP
KKPAGGHVLAHAATIRLLFRKGKGDTRVCKVYDAPNLAEAEASFQITQGGIADAKD"
CDS join(24778. .25277,25542. .25947,26035. .26295,26369. .26455,
26538. .26675)
/note="gene_id:F5N5.6"
/codon_start=1
/evidence=not_experimental
/product="ATP sulfurylase/APS kinase"
/protein_id="BAB03034.1"
/db_xref="GI:11994718"
/translation="MASMAAVLSKTPFLSQPLTKSSPNSDLPFAAVSFPSKSLRRRVG
SIRAGLIAPDGGKLVELIVEEPKRREKKHEAADLPRVELTAIDLQWMHVLSEGWASPL
GGFMRESEFLQTLHFNSLRLDDGSVVNMSVPIVLAIDDEQKARIGESTRVALFNSDGN
PVAILSDIEIYKHPKEERIARTWGTTAPGLPYVDEAITNAGNWLIGGDLEVLEPVKYN
DGLDRFRLSPAELRKELEKRNADAVFAFQLRNPVHNGHALLMTDTRRRLLEMGYKNPI
LLLHPLGGFTKADDVPLDWRMKQHEKVLEDGVLDPETTVVSIFPSPMHYAGPTEVQWH
AKARINAGANFYIVGRDPAGMGHPVEKRDLYDADHGKKVLSMAPGLERLNILPFRVAA
YDKTQGKMAFFDPSRPQDFLFISGTKMRTLAKNNENPPDGFMCPGGWKVLVDYYESLT
PAGNGRLPEVVPV"
CDS 27637. .28110
/note="gene_id:F5N5.7"
/codon_start=1
/evidence=not_experimental
/product="RNA polymerase II fifth largest subunit-like protein"
/protein_id="BAB03035.1"
/db_xref="GI:11994719"
/translation="MDTGLMLQRAIVIRLLEAFSKEKATKDLGYLITPTILENIGEGK
IKEQTGEIQFPVVFNGICFKMFKGEIVHGVVHKVHKTGVFLKSGPYEIIYLSHMKMPG
YEFIPGENPFFMNQYMSRIQIGARVRFVVLDTEWREAEKDFMALASIDGDNLGPF"
CDS complement(28276. .31329)
/note="gene_id:F5N5.8"
/codon_start=1
/evidence=not_experimental
/product="Ca2+-transporting ATPase-like protein"
/protein_id="BAB03036.1"
/db_xref="GI:11994720"
/translation="MRRNVSDHAEKKDKVGVEVLLELPKTLSKSNKKWQLALIKLYCS
RTLLNCAKHAIRKPGLFPRSLSYTAIDLDHHHGDDHFKIDTETLNDLVKNKNQEKLES
LGGPNGLVSALKSNTRLGINEEGDEIQRRRSTFGSNTYTRQPSKGLFHFVVEAFKDLT
ILILLGCATLSLGFGIKEHGLKEGWYDGGSIFVAVFLVVAVSAVSNFRQNRQFDKLSK
VSSNIKIDVVRNGRRQEISIFDIVVGDIVCLNIGDQVPADGVFVEGHLLHVDESSMTG
ESDHVEVSLTGNTFLFSGTKIADGFGKMAVTSVGMNTAWGQMMSHISRDTNEQTPLQS
RLDKLTSSIGKVGLLVAFLVLLVLLIRYFTGTTKDESGNREYNGKTTKSDEIVNAVVK
MVAAAVTIIVVAIPEGLPLAVTLTLAYSMKRMMKDNAMVRKLSACETMGSATVICTDK
TGTLTLNQMKVTDFWFGLESGKASSVSQRVVELFHQGVAMNTTGSVFKAKAGTEYEFS
GSPTEKAILSWAVEELEMGMEKVIEEHDVVHVEGFNSEKKRSGVLMKKKGVNTENNVV
HWKGAAEKILAMCSTFCDGSGVVREMKEDDKIQFEKIIQSMAAKSLRCIAFAYSEDNE
DNKKLKEEKLSLLGIIGIKDPCRPGVKKAVEDCQFAGVNIKMITGDNIFTARAIAVEC
GILTPEDEMNSEAVLEGEKFRNYTQEERLEKVERIKVMARSSPFDKLLMVKCLKELGH
VVAVTGDGTNDAPALKEADIGLSMGIQGTEVAKESSDIVILDDNFASVATVLKWGRCV
YNNIQKFIQFQLTVNVAALVINFVAAVSAGDVPLTAVQLLWVNLIMDTLGALALATEK
PTNDLMKKKPIGRVAPLITNIMWRNLLAQAFYQISVLLVLQFRGRSIFNVTEKVKNTL
IFNTFVLCQVFNEFNARSLEKKNVFKGLHKNRLFIGIIVVTVVLQVVMVEFLKRFADT
ERLNLGQWGVCIAIAAASWPIGWLVKSVPVPERHFFSYLKWKKRS"
CDS complement(34654. .35352)
/note="unnamed protein product; contains similarity to peptidyl-prolyl cis-trans isomerase
gene_id:F5N5.9"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB03037.1"
/db_xref="GI:11994721"
/translation="MANPKVFFDLTVDGKPAGRIVIELFADLTPRTAENFRGLCTGER
GIGKCGKPIHYKGSTFDHIVPDLMWCGGDIIFENEPIHSEELDDEYFILNHEDGPGII
SMADSNGSQFQIHMKDYGLQVDGDHVVIGKVVEGLDLMRNIEKEVITTTRTPSKPVV
IADCGELSDYRSERCYLMKNIEKEVIIKTAKDNKPVVIADCGGLSDDRSERYYLINIV
VACMVLMCFWSWFV"
CDS complement(join(36227. .36364,37307. .37387,37474. .37776))
/note="unnamed protein product; contains similarity to calmodulin
gene_id:F5N5.10"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB03038.1"
/db_xref="GI:11994722"
/translation="MEEIQQQQQQQQQQQQQQQQQQQQQELTQEQIMEFKEAFCLFD
KDGDGCITADELATVIRSLDQNPTEQELQDMITEIDSDGNGTIEFSEFLNLMANQLQE
TDADEELKEAFKVFDKDQNGYISASELRHVMINLGEKLTDEEVDQMIKEADLDGDGQV

```
Alignment Scores:
Pred. No.:              1.42e+04        Length:         71327
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x AP001300 (1-71327)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      54914 CACAATCACAATCACAAT 54897

RESULT 481
DMBN4L24/c
LOCUS       DMBN4L24               71813 bp    DNA     linear   INV 10-OCT-1999
DEFINITION  Drosophila melanogaster clone BACN4L24.
ACCESSION   AL121805
VERSION     AL121805.1  GI:6018875
KEYWORDS    .
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 71813)
  AUTHORS   Mottier, Cadieu, Dreano, Lelaure and Galibert,F.
  TITLE     Sequencing the distal X chromosome of Drosophila melanogaster
  JOURNAL   Unpublished
  REMARK    UPR41 CNRS Recombinaisons Genetiques, Universite de Rennes 1 2 av
            du Pr Leon Bernard,Faculte de Medecine, 35043 RENNES Cedex, France.
REFERENCE   2  (bases 1 to 71813)
  AUTHORS   Benos,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-OCT-1999) European Drosophila Genome Sequencing
            Consortium
COMMENT     Sequence submitted by Takis Benos, EMBL Outstation - The EBI,
            Hinxton Hall, Hinxton, Cambridge, CB10 1SD, U.K.
            E-mail: benos@ebi.ac.uk on behalf of the European Drosophila Genome
            Sequencing Consortium. For further information see the European
            Drosophila Genome Sequencing Consortium's web site:
            http://edgp.ebi.ac.uk/.
            This is the finished sequence of clone BACN4L24.  The CDSs
            presented here are only preliminary predictions based on automatic
            analysis. Analysis of the sequence is in progress.
            Clone=BACN4L24; Contig ID=0; Length=71813; Status=Finished.
FEATURES             Location/Qualifiers
     source          1. .71813
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN4L24"
ORIGIN

Alignment Scores:
Pred. No.:              1.43e+04        Length:         71813
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-858-332G-28 (1-6) x DMBN4L24 (1-71813)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      18295 CACAATCACAATCACAAT 18278

RESULT 482
CQ598493/c
LOCUS       CQ598493               71962 bp    DNA     linear   PAT 02-FEB-2004
DEFINITION  Sequence 26251 from Patent WO0171042.
ACCESSION   CQ598493
VERSION     CQ598493.1  GI:41654308
KEYWORDS    .
SOURCE      Drosophila sp.
  ORGANISM  Drosophila sp.
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1
  AUTHORS   Venter,J.C., Adams,M., Li,P.W. and Myers,E.W.
  TITLE     Detection kits, such as nucleic acid arrays, for detecting the
            expression of 10,000 or more Drosophila genes and uses thereof
  JOURNAL   Patent: WO 0171042-A 26251 27-SEP-2001;
            PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1. .71962
                     /organism="Drosophila sp."
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:7242"
ORIGIN

Alignment Scores:
Pred. No.:              1.44e+04        Length:         71962
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x CQ598493 (1-71962)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      21833 CATAATCATAATCATAAT 21816

RESULT 483
AL607071
LOCUS       AL607071               72006 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP11-121B21 on chromosome 6, complete
            sequence.
ACCESSION   AL607071
VERSION     AL607071.7  GI:16944255
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 72006)
  AUTHORS   Griffiths,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 15, 2001 this sequence version replaced gi:15982163.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP11-121B21 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            ---------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
```

```
            Contact: vega@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1. .72006
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-121B21"
                     /clone_lib="RPCI-11.1"
     misc_feature    2000
                     /note="Clone_right_end: RP11-436D23"
     misc_feature    70007
                     /note="Clone_left_end: RP3-453D15"
ORIGIN

Alignment Scores:
Pred. No.:                1.44e+04        Length:         72006
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       8               Gaps:           0

US-09-858-332G-28 (1-6) x AL607071 (1-72006)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      31101 CATAACCATAACCACAAC 31118

RESULT 484
AC004713/c
LOCUS       AC004713               72327 bp    DNA     linear   INV 31-OCT-1998
DEFINITION  Drosophila melanogaster, chromosome 3L, region 62B8-62B11, P1 clone
            DS04710, complete sequence.
ACCESSION   AC004713
VERSION     AC004713.1  GI:3818346
KEYWORDS    HTG.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 72327)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Punch,E., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Twomey,B., Wan,K.H., Weinburg,T., Zhang,R.,
            Zieran,L.L. and Rubin,G.M.
  TITLE     Sequencing of Dras3 gene
  JOURNAL   Unpublished (1998)
REFERENCE   2  (bases 1 to 72327)
  AUTHORS   Celniker,S.E., George,R.A., Galle,R., Svirskas,R.R., Hoskins,R.A.,
            Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C.,
            Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
            Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
            Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E.,
            Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R.,
            Zieran,L.L. and Kimmel,B.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAY-1998) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Oct 31, 1998 this sequence version replaced gi:3805737.
            Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory, MS 64-121
            Berkeley, CA 94720
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu.
            P1 library location: 50-6.
FEATURES             Location/Qualifiers
     source          1. .72327
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /strain="y2; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="3L"
                     /map="62B8-62B11"
                     /clone="P1 DS04710 (D227)"
                     /clone_lib="P1 library, partial Sau3A in pNS582tet14Ad10"
                     /note="DS04710 (D227) is a partial bridge clone which
                     minimally overlaps DS04808 (D228) at bp 1 and extends to
                     P1 end at bp 72,327."
ORIGIN

Alignment Scores:
Pred. No.:                1.44e+04        Length:         72327
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       2               Gaps:           0

US-09-858-332G-28 (1-6) x AC004713 (1-72327)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1872 CATAATCATAATCATAAT 1855

RESULT 485
AC131257
LOCUS       AC131257               72409 bp    DNA     linear   HTG 26-SEP-2002
DEFINITION  Homo sapiens chromosome 15 clone RP13-423F3 map 15, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC131257
VERSION     AC131257.3  GI:23322701
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 72409)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP13-423F3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 72409)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
```

Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (20-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 72409)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (26-SEP-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Sep 26, 2002 this sequence version replaced gi:22657542.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-------------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-------------- Project Information
Center project name: L28022
Center clone name: 423_F_3
--------------.
* NOTE: This record contains 86 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

*	1	713: contig of 713 bp in length
*	714	813: gap of 100 bp
*	814	1543: contig of 730 bp in length
*	1544	1643: gap of 100 bp
*	1644	2398: contig of 755 bp in length
*	2399	2498: gap of 100 bp
*	2499	3220: contig of 722 bp in length
*	3221	3320: gap of 100 bp
*	3321	4071: contig of 751 bp in length
*	4072	4171: gap of 100 bp
*	4172	4928: contig of 757 bp in length
*	4929	5028: gap of 100 bp
*	5029	5783: contig of 755 bp in length
*	5784	5883: gap of 100 bp
*	5884	6640: contig of 757 bp in length
*	6641	6740: gap of 100 bp
*	6741	7484: contig of 744 bp in length
*	7485	7584: gap of 100 bp
*	7585	8329: contig of 745 bp in length
*	8330	8429: gap of 100 bp
*	8430	9147: contig of 718 bp in length
*	9148	9247: gap of 100 bp
*	9248	9992: contig of 745 bp in length
*	9993	10092: gap of 100 bp
*	10093	10821: contig of 729 bp in length
*	10822	10921: gap of 100 bp
*	10922	11666: contig of 745 bp in length
*	11667	11766: gap of 100 bp
*	11767	12497: contig of 731 bp in length
*	12498	12597: gap of 100 bp
*	12598	13321: contig of 724 bp in length
*	13322	13421: gap of 100 bp
*	13422	14166: contig of 745 bp in length
*	14167	14266: gap of 100 bp
*	14267	14999: contig of 733 bp in length
*	15000	15099: gap of 100 bp
*	15100	15847: contig of 748 bp in length
*	15848	15947: gap of 100 bp
*	15948	16693: contig of 746 bp in length
*	16694	16793: gap of 100 bp
*	16794	17536: contig of 743 bp in length
*	17537	17636: gap of 100 bp
*	17637	18383: contig of 747 bp in length
*	18384	18483: gap of 100 bp
*	18484	19222: contig of 739 bp in length
*	19223	19322: gap of 100 bp
*	19323	20050: contig of 728 bp in length
*	20051	20150: gap of 100 bp
*	20151	20895: contig of 745 bp in length
*	20896	20995: gap of 100 bp
*	20996	21747: contig of 752 bp in length
*	21748	21847: gap of 100 bp
*	21848	22595: contig of 748 bp in length
*	22596	22695: gap of 100 bp
*	22696	23446: contig of 751 bp in length
*	23447	23546: gap of 100 bp
*	23547	24299: contig of 753 bp in length
*	24300	24399: gap of 100 bp
*	24400	25155: contig of 756 bp in length
*	25156	25255: gap of 100 bp
*	25256	25994: contig of 739 bp in length
*	25995	26094: gap of 100 bp
*	26095	26856: contig of 762 bp in length
*	26857	26956: gap of 100 bp
*	26957	27697: contig of 741 bp in length
*	27698	27797: gap of 100 bp
*	27798	28541: contig of 744 bp in length
*	28542	28641: gap of 100 bp
*	28642	29384: contig of 743 bp in length
*	29385	29484: gap of 100 bp
*	29485	30222: contig of 738 bp in length
*	30223	30322: gap of 100 bp
*	30323	31069: contig of 747 bp in length
*	31070	31169: gap of 100 bp
*	31170	31903: contig of 734 bp in length
*	31904	32003: gap of 100 bp
*	32004	32747: contig of 744 bp in length
*	32748	32847: gap of 100 bp
*	32848	33600: contig of 753 bp in length
*	33601	33700: gap of 100 bp
*	33701	34456: contig of 756 bp in length
*	34457	34556: gap of 100 bp
*	34557	35310: contig of 754 bp in length
*	35311	35410: gap of 100 bp
*	35411	36158: contig of 748 bp in length
*	36159	36258: gap of 100 bp
*	36259	36991: contig of 733 bp in length
*	36992	37091: gap of 100 bp
*	37092	37834: contig of 743 bp in length
*	37835	37934: gap of 100 bp
*	37935	38672: contig of 738 bp in length
*	38673	38772: gap of 100 bp
*	38773	39518: contig of 746 bp in length
*	39519	39618: gap of 100 bp

```
            *    39619    40343: contig of 725 bp in length
            *    40344    40443: gap of 100 bp
            *    40444    41189: contig of 746 bp in length
            *    41190    41289: gap of 100 bp
            *    41290    42042: contig of 753 bp in length
            *    42043    42142: gap of 100 bp
            *    42143    42898: contig of 756 bp in length
            *    42899    42998: gap of 100 bp
            *    42999    43759: contig of 761 bp in length
            *    43760    43859: gap of 100 bp
            *    43860    44616: contig of 757 bp in length
            *    44617    44716: gap of 100 bp
            *    44717    45466: contig of 750 bp in length
            *    45467    45566: gap of 100 bp
            *    45567    46312: contig of 746 bp in length
            *    46313    46412: gap of 100 bp
            *    46413    47151: contig of 739 bp in length
            *    47152    47251: gap of 100 bp
            *    47252    48003: contig of 752 bp in length
            *    48004    48103: gap of 100 bp
            *    48104    48861: contig of 758 bp in length
            *    48862    48961: gap of 100 bp
            *    48962    49715: contig of 754 bp in length

Alignment Scores:
Pred. No.:                1.45e+04        Length:         72409
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       14              Gaps:           0

US-09-858-332G-28 (1-6) x AC131257 (1-72409)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      10555 CATAATCATAACCATAAC 10572

RESULT 486
CT005260_7
WPCOMMENT
  Sequence split into 8 fragments  LOCUS CT005260 Accession CT005260
     Fragment Name      Begin     End
     CT005260_0             1    110000
     CT005260_1        100001    210000
     CT005260_2        200001    310000
     CT005260_3        300001    410000
     CT005260_4        400001    510000
     CT005260_5        500001    610000
     CT005260_6        600001    710000
     CT005260_7        700001    772974
Continuation (8 of 8) of CT005260 from base 700001 (CT005260 Leishmania major strain Fri

Alignment Scores:
Pred. No.:                1.46e+04        Length:         72974
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       14              Gaps:           0

US-09-858-332G-28 (1-6) x CT005260_7 (1-72974)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      22215 CACAACCACAACCACAAC 22232

RESULT 487
OSIG00033/c
LOCUS       OSIG00033             73187 bp    DNA     linear   HTG 01-AUG-2002
DEFINITION  Oryza sativa (indica cultivar-group) chromosome 4 clone H0315F07,
            *** SEQUENCING IN PROGRESS ***.
ACCESSION   AL732338
VERSION     AL732338.1  GI:22091843
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Han,B., Zhou,B., Feng,Q., Huang,Y.C., Li,Y., Zhu,J.J., Zhao,Q.,
            Hu,X., Liu,Y.L., Mu,J., Yu,Z., Chen,L., Fan,D.L., Weng,Q.J.,
            Zhang,L., Lu,Y.Q., Yu,S.L., Liu,X.H., Lu,T.T., Zhang,Y.J., Lu,Y.,
            Li,C., Li,T., Zhang,Y., Hu,H., Jia,P.X., Qian,Y.M., Ying,K.,
            Chen,Z.H., Hao,P., Zhang,L., Wu,M., Zhang,R.Q., Guan,J.P. and
            Hong,G.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAY-2002) Han Bin, National Center for Gene Research,
            Chinese Academy of sciences, 500# Cao Bao Road, Shanghai 200233,
            CHINA. E-mail enquiries: bhan@ncgr.ac.cn. Clone requests:
            bhan@ncgr.ac.cn
  REMARK    Oryza sativa indica (guangluai4) genomic DNA, chromosome 4, BAC
            clone: H0315F07.
COMMENT     Web site: http://www.ncgr.ac.cn
            -------------- Summary Statistics
            Assembly program: phrap
            --------------
            This is a complete sequence.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .73187
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /variety="GuangLuAi4"
                     /sub_species="indica"
                     /db_xref="taxon:39946"
                     /chromosome="4"
                     /clone="H0315F07"
                     /clone_lib="BAC-HindIII"
ORIGIN

Alignment Scores:
Pred. No.:                1.46e+04        Length:         73187
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       14              Gaps:           0

US-09-858-332G-28 (1-6) x OSIG00033 (1-73187)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      61153 CACAACCACAACCACAAC 61136

RESULT 488
AC067962/c
LOCUS       AC067962              73565 bp    DNA     linear   PRI 15-APR-2005
DEFINITION  Homo sapiens BAC clone RP11-48K7 from 2, complete sequence.
ACCESSION   AC067962
VERSION     AC067962.6  GI:18042343
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 73565)
  AUTHORS   Harris,A. and Haakenson,W.
  TITLE     The sequence of Homo sapiens BAC clone RP11-48K7
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 73565)
```

```
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3  (bases 1 to 73565)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JAN-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 73565)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JAN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   5  (bases 1 to 73565)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-APR-2005) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jan 3, 2002 this sequence version replaced gi:10242446.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0048K07
            --------------.

            NOTICE:

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR:  pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-505O20, 2000 bp overlap.
            Actual start of this clone is at base position 105068 of
            RP11-505O20; actual end is at base position 73565 of RP11-48K7.
FEATURES             Location/Qualifiers
     source          1. .73565
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="RP11-48K7"
                     /clone_lib="RPCI-11"
     misc_feature    70383. .71338
                     /note="CpG_island (%GC=67.2, o/e=0.72, #CpGs=71)"
     misc_feature    71709. .71970
                     /note="CpG_island (%GC=61.1, o/e=0.81, #CpGs=21)"
ORIGIN

Alignment Scores:
Pred. No.:              1.47e+04        Length:         73565
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     8               Gaps:           0

US-09-858-332G-28 (1-6) x AC067962 (1-73565)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       7613 CACAATCACAACCACAAT 7596

RESULT 489
HS1071N3
LOCUS       HS1071N3               74037 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP5-1071N3 on chromosome 1p36.2
            Contains the 5' end of the ECE1 gene for endothelin converting
            enzyme 1, a novel gene and a CpG island, complete sequence.
ACCESSION   AL031728
VERSION     AL031728.12  GI:4972242
KEYWORDS    HTG; CpG island; ECE1.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 74037)
  AUTHORS   Davidson,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 2, 1999 this sequence version replaced gi:4938270.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP5-1071N3 is from the library RPCI-5 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1. .74037
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="RZPD:RPCIP704N031071"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="p36.2"
                     /clone="RP5-1071N3"
```

```
                /clone_lib="RPCI-5"
misc_feature    100
                /note="Clone_right_end: RP3-329E20"
misc_feature    8156. .8393
                /note="Single clone region. Sequence from reads from a
                short insert library derived from a single pUC clone. Part
                of this sequence is confirmed by X91922."
gene            join(complement(63562. .63690),complement(8256. .8342),
                complement(AL031005.2:107183. .107324),
                complement(AL031005.2:100691. .100903),
                complement(AL031005.2:88263. .88384),
                complement(AL031005.2:86685. .86831),
                complement(AL031005.2:85517. .85582),
                complement(AL031005.2:83939. .84130),
                complement(AL031005.2:75213. .75355),
                complement(AL031005.2:72981. .73095),
                complement(AL031005.2:66126. .66236),
                complement(AL031005.2:64738. .64836),
                complement(AL031005.2:63842. .63919),
                complement(AL031005.2:61550. .61653),
                complement(AL031005.2:55923. .56033),
                complement(AL031005.2:55151. .55218),
                complement(AL031005.2:53242. .53432),
                complement(AL031005.2:49739. .49834),
                complement(AL031005.2:47574. .48123))
                /gene="ECE1"
                /locus_tag="RP3-329E20.1-001"
mRNA            join(complement(63562. .63690),complement(8256. .8342),
                complement(AL031005.2:107183. .107324),
                complement(AL031005.2:100691. .100903),
                complement(AL031005.2:88263. .88384),
                complement(AL031005.2:86685. .86831),
                complement(AL031005.2:85517. .85582),
                complement(AL031005.2:83939. .84130),
                complement(AL031005.2:75213. .75355),
                complement(AL031005.2:72981. .73095),
                complement(AL031005.2:66126. .66236),
                complement(AL031005.2:64738. .64836),
                complement(AL031005.2:63842. .63919),
                complement(AL031005.2:61550. .61653),
                complement(AL031005.2:55923. .56033),
                complement(AL031005.2:55151. .55218),
                complement(AL031005.2:53242. .53432),
                complement(AL031005.2:49739. .49834),
                complement(AL031005.2:47574. .48123))
                /gene="ECE1"
                /locus_tag="RP3-329E20.1-001"
                /product="endothelin converting enzyme 1"
                /note="match: ESTs: Em:AA553878.1 Em:AI133414.1
                Em:AI270018.1 Em:AI417789.1 Em:AL554172.2 Em:AL576971.2
                Em:AU099212.1 Em:AU131871.1 Em:AU133165.1 Em:BE535389.1
                Em:BE563471.1 Em:BE812073.1 Em:BE867023.1 Em:BE871194.1
                Em:BF344246.1 Em:BF939283.1 Em:BF969352.1 Em:BG023787.1
                Em:BG283366.1 Em:BG326696.1 Em:BG331831.1 Em:BG681313.1
                Em:BG760802.1 Em:BG821605.1 Em:BG827870.1 Em:BI869244.1
                Em:BM043900.1 Em:BM051302.1 Em:BM669522.1 Em:BM673817.1
                Em:BM684594.1 Em:BM922007.1 Em:BM931139.1 Em:BM976164.1
                Em:BM977396.1 Em:BQ005963.1 Em:BQ213331.1 Em:BQ219825.1
                Em:BQ224802.1 Em:BQ324019.1 Em:BQ420367.1 Em:BQ423018.1
                Em:BQ679953.1 Em:BQ684918.1 Em:BQ722674.1 Em:BQ883990.1
                Em:BU195928.1 Em:BU430876.1 Em:BU501029.1 Em:BU538295.1
                Em:BU629362.1 Em:BU733112.1 Em:BU845418.1 Em:BU931731.1
                Em:BX328039.1 Em:BX475348.1 Em:CA310174.1 Em:CA419827.1
                Em:CA449233.1 Em:CA489517.1 Em:CB853448.1 Em:CB989367.1
                Em:CD244021.1 Em:CD244093.1 Em:CD364951.1 Em:CD514948.1
                Em:CD644543.1 Em:CD652529.1 Em:CF125824.1
                match: cDNAs: Em:AB031742.1"
CDS             join(complement(63562. .63564),complement(8256. .8342),
                complement(AL031005.2:107183. .107324),
                complement(AL031005.2:100691. .100903),
                complement(AL031005.2:88263. .88384),
                complement(AL031005.2:86685. .86831),
                complement(AL031005.2:85517. .85582),
                complement(AL031005.2:83939. .84130),
                complement(AL031005.2:75213. .75355),
                complement(AL031005.2:72981. .73095),
                complement(AL031005.2:66126. .66236),
                complement(AL031005.2:64738. .64836),
                complement(AL031005.2:63842. .63919),
                complement(AL031005.2:61550. .61653),
                complement(AL031005.2:55923. .56033),
                complement(AL031005.2:55151. .55218),
                complement(AL031005.2:53242. .53432),
                complement(AL031005.2:49739. .49834),
                complement(AL031005.2:47947. .48123))
                /gene="ECE1"
                /locus_tag="RP3-329E20.1-001"
                /standard_name="OTTHUMP00000002724"
                /codon_start=1
                /product="endothelin converting enzyme 1"
                /protein_id="CAI19064.1"
                /db_xref="GI:56203021"
                /translation="MMSTYKRATLDEEDLVDSLSEGDAYPNGLQVNFHSPRSGQRCWA
                ARTQVEKRLVVLVVLLAAGLVACLAALGIQYQTRSPSVCLSEACVSVTSSILSSMDPT
                VDPCHDFFSYACGGWIKANPVPDGHSRWGTFSNLWEHNQAIIKHLLENSTASVSEAER
                KAQVYYRACMNETRIEELRAKPLMELIERLGGWNITGPWAKDNFQDTLQVVTAHYRTS
                PFFSVYVSADSKNSNSNVIQVDQSGLGLPSRDYYLNKTENEKVLTGYLNYMVQLGKLL
                GGGDEEAIRPQMQQILDFETALANITIPQEKRRDEELIYHKVTAAELQTLAPAINWLP
                FLNTIFYPVEINESEPIVVYDKEYLEQISTLINTTDRCLLNNYMIWNLVRKTSSFLDQ
                RFQDADEKFMEVMYGTKKTCLPRWKFCVSDTENNLGFALGPMFVKATFAEDSKSIATE
                IILEIKKAFEESLSTLKWMDEETRKSAKEKADAIYNMIGYPNFIMDPKELDKVFNDYT
                AVPDLYFENAMRFFNFSWRVTADQLRKAPNRDQWSMTPPMVNAYYSPTKNEIVFPAGI
                LQAPFYTRSSPKALNFGGIGVVVGHELTHAFDDQGREYDKDGNLRPWWKNSSVEAFKR
                QTECMVEQYSNYSVNGEPVNGRHTLGENIADNGGLKAAYRAYQNWVKKNGAEHSLPTL
                GLTNNQLFFLGFAQVWCSVRTPESSHEGLITDPHSPSRFRVIGSLSNSKEFSEHFRCP
                PGSPMNPPHKCEVW"
gene            join(complement(63123. .63323),complement(8256. .8342),
                complement(AL031005.2:107182. .107324),
                complement(AL031005.2:100827. .100903))
                /gene="ECE1"
                /locus_tag="RP3-329E20.1-004"
mRNA            join(complement(63123. .63323),complement(8256. .8342),
                complement(AL031005.2:107182. .107324),
                complement(AL031005.2:100827. .100903))
                /gene="ECE1"
                /locus_tag="RP3-329E20.1-004"
                /product="endothelin converting enzyme 1"
                /note="match: ESTs: Em:AI522130.1"
gene            join(complement(8550. .8600),complement(8256. .8342),
                complement(AL031005.2:107183. .107324),
                complement(AL031005.2:100691. .100903),
                complement(AL031005.2:88263. .88384),
                complement(AL031005.2:86685. .86831),
                complement(AL031005.2:85517. .85582),
                complement(AL031005.2:83939. .84130),
                complement(AL031005.2:75213. .75355),
                complement(AL031005.2:72981. .73095),
                complement(AL031005.2:66126. .66236),
                complement(AL031005.2:64738. .64836),
                complement(AL031005.2:63842. .63919),
                complement(AL031005.2:61550. .61653),
                complement(AL031005.2:55923. .56033),
                complement(AL031005.2:55151. .55218),
                complement(AL031005.2:53242. .53432),
                complement(AL031005.2:49739. .49834),
                complement(AL031005.2:47851. .48123))
                /gene="ECE1"
                /locus_tag="RP3-329E20.1-002"
mRNA            join(complement(8550. .8600),complement(8256. .8342),
                complement(AL031005.2:107183. .107324),
                complement(AL031005.2:100691. .100903),
                complement(AL031005.2:88263. .88384),
                complement(AL031005.2:86685. .86831),
                complement(AL031005.2:85517. .85582),
                complement(AL031005.2:83939. .84130),
                complement(AL031005.2:75213. .75355),
```

```
Alignment Scores:
Pred. No.:                1.48e+04        Length:         74037
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       8               Gaps:           0

US-09-858-332G-28 (1-6) x HS1071N3 (1-74037)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      37545 CATAATCATAATCATAAT 37562

RESULT 490
AF260012
LOCUS       AF260012               74509 bp    DNA     linear   HTG 26-JUL-2002
DEFINITION  Homo sapiens chromosome 8 clone RP3-491L6 map 8q2, WORKING DRAFT
            SEQUENCE, 4 unordered pieces.
ACCESSION   AF260012
VERSION     AF260012.3  GI:14280186
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_CANCELLED.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 74509)
  AUTHORS   Schilhabel,M.B., Baumgart,C., Blechschmidt,K., Dette,M., Jahn,N.,
            Lehmann,R., Menzel,U., Polley,A., Reichwald,K., Schudy,A.,
            Siddiqui,R., Taudien,S., Wen,G., Rosenthal,A. and Platzer,M.
  TITLE     Chromosome 8 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 74509)
  AUTHORS   Polley,A., Wen,G., Baumgart,C., Dette,M., Jahn,N., Schilhabel,M.,
            Menzel,U. and Rosenthal,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2000) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
COMMENT     On Jun 2, 2001 this sequence version replaced gi:8151846.
            -------------- Genome Center
                 Center: Insitute of Molecular Biotechnoloy
                 Center code: IMB
                 Web site: http://genome.imb-jena.de/
                 Contact: gscj-submit@genome.imb-jena.de
             -------------- Project Information
                 Center project name: H238
                 Center clone name: RP3-491L6
             -------------- Summary Statistics
                 Sequencing vector: M13; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990329
                 Consensus quality: 69941 bases at least Q40
                 Consensus quality: 71673 bases at least Q30
                 Consensus quality: 73001 bases at least Q20
                 Quality coverage: 5,68 in Q20 bases; sum-of-contigs
             --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3432: contig of 3432 bp in length
            *     3433     3532: gap of unknown length
            *     3533     7836: contig of 4304 bp in length
            *     7837     7936: gap of unknown length
            *     7937    47443: contig of 39507 bp in length
            *    47444    47543: gap of unknown length
            *    47544    74509: contig of 26966 bp in length.
FEATURES             Location/Qualifiers
     source          1. .74509
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8q2"
                     /clone="RP3-491L6"
     gap             3433. .3532
                     /estimated_length=unknown
     gap             7837. .7936
                     /estimated_length=unknown
     gap             47444. .47543
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:                1.49e+04        Length:         74509
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       14              Gaps:           0

US-09-858-332G-28 (1-6) x AF260012 (1-74509)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      60984 CACAATCATAATCATAAT 61001

RESULT 491
AC091545
LOCUS       AC091545               75148 bp    DNA     linear   PRI 04-MAR-2002
DEFINITION  Homo sapiens chromosome 8, clone RP3-491L6, complete sequence.
ACCESSION   AC091545
VERSION     AC091545.4  GI:19073804
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 75148)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP3-491L6
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 75148)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
            Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K.,
            Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
            Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 75148)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
```

```
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 75148)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 4, 2002 this sequence version replaced gi:18693435.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L12497
                Center clone name: 491_L_6
            --------------.
FEATURES            Location/Qualifiers
     source         1. .75148
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /chromosome="8"
                    /map="8"
                    /clone="RP3-491L6"
                    /clone_lib="RPCI Human PAC library 3"
     unsure         1845. .1851
                    /note="<30 qual SNGL region"
     unsure         1911. .1915
                    /note="<30 qual SNGL region"
     repeat_region  complement(1984. .2083)
                    /rpt_family="MER2"
     repeat_region  complement(2191. .2642)
                    /rpt_family="L1MB4"
     repeat_region  3000. .3066
                    /rpt_family="MIR"
     repeat_region  complement(3165. .3343)
                    /rpt_family="MIR"
     repeat_region  3821. .3971
                    /rpt_family="MIR"
     repeat_region  10524. .10758
                    /rpt_family="MIR"
     repeat_region  14452. .14749
                    /rpt_family="AluSx"
     repeat_region  complement(15073. .15367)
                    /rpt_family="AluJb"
     repeat_region  complement(16235. .16551)
                    /rpt_family="L2"
     repeat_region  complement(16906. .17212)
                    /rpt_family="AluSx"
     repeat_region  complement(17607. .17718)
                    /rpt_family="L3"
     repeat_region  complement(17954. .18239)
                    /rpt_family="L3"
     repeat_region  19486. .19514
                    /rpt_family="(CAAAA)n"
     repeat_region  complement(19634. .20081)
                    /rpt_family="L1ME1"
     repeat_region  complement(20211. .20513)
                    /rpt_family="AluJo"
     repeat_region  complement(20598. .20901)
                    /rpt_family="AluJo"
     repeat_region  complement(20952. .21209)
                    /rpt_family="L1MA3"
     repeat_region  22781. .22820
                    /rpt_family="AT_rich"
     repeat_region  complement(24824. .25124)
                    /rpt_family="AluY"
     repeat_region  complement(25280. .25337)
                    /rpt_family="MIR"
     repeat_region  complement(25546. .25681)
                    /rpt_family="MIR"
     repeat_region  complement(25692. .25827)
                    /rpt_family="L2"
     repeat_region  complement(25863. .26330)
                    /rpt_family="L2"
     repeat_region  complement(27189. .27476)
                    /rpt_family="AluSx"
     repeat_region  30875. .30981
                    /rpt_family="MIR"
     repeat_region  30982. .31286
                    /rpt_family="AluSx"
     repeat_region  31287. .31340
                    /rpt_family="MIR"
     repeat_region  32152. .32290
                    /rpt_family="MIR"
     repeat_region  33175. .33550
                    /rpt_family="MLT1B"
     repeat_region  33710. .33794
                    /rpt_family="MIR"
     repeat_region  33995. .34103
                    /rpt_family="MIR3"
     repeat_region  complement(35542. .36082)
                    /rpt_family="L1ME1"
     repeat_region  complement(36083. .36380)
                    /rpt_family="AluSx"
     repeat_region  complement(36381. .36542)
                    /rpt_family="L1ME1"
     repeat_region  complement(36556. .36722)
                    /rpt_family="AluJb"
```

```
     repeat_region   36737. .36799
                     /rpt_family="CT-rich"
     repeat_region   complement(36800. .36932)
                     /rpt_family="FLAM_C"
     repeat_region   complement(36935. .37289)
                     /rpt_family="L1ME1"
     repeat_region   complement(37287. .37588)
                     /rpt_family="L1ME1"
     repeat_region   complement(37700. .38183)
                     /rpt_family="MLT1D"
     repeat_region   complement(38322. .38632)
                     /rpt_family="AluSq"
     repeat_region   complement(38729. .38975)

Alignment Scores:
Pred. No.:              1.5e+04       Length:         75148
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     8             Gaps:           0

US-09-858-332G-28 (1-6) x AC091545 (1-75148)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      61623 CACAATCATAATCATAAT 61640

RESULT 492
AC101416/c
LOCUS       AC101416               75429 bp    DNA     linear   HTG 23-NOV-2001
DEFINITION  Mus musculus clone RP23-121K13, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC101416
VERSION     AC101416.1  GI:17060191
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 75429)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-121K13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 75429)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
-------------- Genome Center
    Center: Whitehead Institute/ MIT Center for Genome Research
    Center code: WIBR
    Web site: http://www-seq.wi.mit.edu
    Contact: sequence_submissions@genome.wi.mit.edu
-------------- Project Information
    Center project name: L16650
    Center clone name: 121_K_13
--------------.
* NOTE: This record contains 90 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*        1      709: contig of 709 bp in length
*      710      809: gap of 100 bp
*      810     1538: contig of 729 bp in length
*     1539     1638: gap of 100 bp
*     1639     2365: contig of 727 bp in length
*     2366     2465: gap of 100 bp
*     2466     3192: contig of 727 bp in length
*     3193     3292: gap of 100 bp
*     3293     4039: contig of 747 bp in length
*     4040     4139: gap of 100 bp
*     4140     4879: contig of 740 bp in length
*     4880     4979: gap of 100 bp
*     4980     5690: contig of 711 bp in length
*     5691     5790: gap of 100 bp
*     5791     6483: contig of 693 bp in length
*     6484     6583: gap of 100 bp
*     6584     7318: contig of 735 bp in length
*     7319     7418: gap of 100 bp
*     7419     8168: contig of 750 bp in length
*     8169     8268: gap of 100 bp
*     8269     8992: contig of 724 bp in length
*     8993     9092: gap of 100 bp
*     9093     9814: contig of 722 bp in length
*     9815     9914: gap of 100 bp
*     9915    10666: contig of 752 bp in length
*    10667    10766: gap of 100 bp
*    10767    11512: contig of 746 bp in length
*    11513    11612: gap of 100 bp
*    11613    12356: contig of 744 bp in length
*    12357    12456: gap of 100 bp
*    12457    13155: contig of 699 bp in length
*    13156    13255: gap of 100 bp
*    13256    13975: contig of 720 bp in length
*    13976    14075: gap of 100 bp
*    14076    14846: contig of 771 bp in length
*    14847    14946: gap of 100 bp
*    14947    15674: contig of 728 bp in length
*    15675    15774: gap of 100 bp
*    15775    16538: contig of 764 bp in length
*    16539    16638: gap of 100 bp
*    16639    17365: contig of 727 bp in length
*    17366    17465: gap of 100 bp
*    17466    18200: contig of 735 bp in length
*    18201    18300: gap of 100 bp
*    18301    19039: contig of 739 bp in length
*    19040    19139: gap of 100 bp
*    19140    19877: contig of 738 bp in length
*    19878    19977: gap of 100 bp
*    19978    20724: contig of 747 bp in length
*    20725    20824: gap of 100 bp
*    20825    21578: contig of 754 bp in length
*    21579    21678: gap of 100 bp
*    21679    22430: contig of 752 bp in length
```

```
*    22431    22530: gap of 100 bp
*    22531    23281: contig of 751 bp in length
*    23282    23381: gap of 100 bp
*    23382    24115: contig of 734 bp in length
*    24116    24215: gap of 100 bp
*    24216    24948: contig of 733 bp in length
*    24949    25048: gap of 100 bp
*    25049    25774: contig of 726 bp in length
*    25775    25874: gap of 100 bp
*    25875    26632: contig of 758 bp in length
*    26633    26732: gap of 100 bp
*    26733    27482: contig of 750 bp in length
*    27483    27582: gap of 100 bp
*    27583    28329: contig of 747 bp in length
*    28330    28429: gap of 100 bp
*    28430    29189: contig of 760 bp in length
*    29190    29289: gap of 100 bp
*    29290    30030: contig of 741 bp in length
*    30031    30130: gap of 100 bp
*    30131    30870: contig of 740 bp in length
*    30871    30970: gap of 100 bp
*    30971    31706: contig of 736 bp in length
*    31707    31806: gap of 100 bp
*    31807    32557: contig of 751 bp in length
*    32558    32657: gap of 100 bp
*    32658    33368: contig of 711 bp in length
*    33369    33468: gap of 100 bp
*    33469    34199: contig of 731 bp in length
*    34200    34299: gap of 100 bp
*    34300    35047: contig of 748 bp in length
*    35048    35147: gap of 100 bp
*    35148    35892: contig of 745 bp in length
*    35893    35992: gap of 100 bp
*    35993    36732: contig of 740 bp in length
*    36733    36832: gap of 100 bp
*    36833    37584: contig of 752 bp in length
*    37585    37684: gap of 100 bp
*    37685    38411: contig of 727 bp in length
*    38412    38511: gap of 100 bp
*    38512    39258: contig of 747 bp in length
*    39259    39358: gap of 100 bp
*    39359    40113: contig of 755 bp in length
*    40114    40213: gap of 100 bp
*    40214    40953: contig of 740 bp in length
*    40954    41053: gap of 100 bp
*    41054    41806: contig of 753 bp in length
*    41807    41906: gap of 100 bp
*    41907    42650: contig of 744 bp in length
*    42651    42750: gap of 100 bp
*    42751    43491: contig of 741 bp in length
*    43492    43591: gap of 100 bp
*    43592    44341: contig of 750 bp in length
*    44342    44441: gap of 100 bp
*    44442    45195: contig of 754 bp in length
*    45196    45295: gap of 100 bp
*    45296    46054: contig of 759 bp in length
*    46055    46154: gap of 100 bp
*    46155    46901: contig of 747 bp in length
*    46902    47001: gap of 100 bp
*    47002    47749: contig of 748 bp in length
*    47750    47849: gap of 100 bp
*    47850    48596: contig of 747 bp in length
*    48597    48696: gap of 100 bp
*    48697    49458: contig of 762 bp in length
*    49459    49558: gap of 100 bp
*    49559    50289: contig of 731 bp in length
*    50290    50389: gap of 100 bp
*    50390    51129: contig of 740 bp in length
*    51130    51229: gap of 100 bp
*    51230    51946: contig of 717 bp in length
*    51947    52046: gap of 100 bp
*    52047    52771: contig of 725 bp in length
*    52772    52871: gap of 100 bp
*    52872    53619: contig of 748 bp in length
*    53620    53719: gap of 100 bp
*    53720    54435: contig of 716 bp in length
*    54436    54535: gap of 100 bp
*    54536    55296: contig of 761 bp in length
*    55297    55396: gap of 100 bp
*    55397    56102: contig of 706 bp in length
*    56103    56202: gap of 100 bp
*    56203    56939: contig of 737 bp in length
*    56940    57039: gap of 100 bp
*    57040    57798: contig of 759 bp in length

Alignment Scores:
Pred. No.:                1.5e+04        Length:          75429
Score:                    42.00          Matches:         6
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       14             Gaps:            0

US-09-858-332G-28 (1-6) x AC101416 (1-75429)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      55653 CACAATCATAATCATAAT 55636

RESULT 493
AC149204
LOCUS       AC149204               75741 bp    DNA     linear   PLN 22-MAY-2004
DEFINITION  Medicago truncatula chromosome 7 BAC clone mth2-77n20, complete
            sequence.
ACCESSION   AC149204
VERSION     AC149204.1  GI:47564143
KEYWORDS    HTG.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 75741)
  AUTHORS   Town,C.D., Tallon,L.J., Arbogast,T., Althoff,R., Hine,E.,
            Monaghan,E., Smith,S.A., Utterback,T., Feldblyum,T., Koo,H. and
            Cheung,F.
  TITLE     Medicago truncatula BAC genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 75741)
  AUTHORS   Town,C.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAY-2004) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .75741
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3880"
                     /chromosome="7"
                     /clone="mth2-77n20"
ORIGIN

Alignment Scores:
Pred. No.:                1.51e+04       Length:          75741
Score:                    42.00          Matches:         6
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       15             Gaps:            0

US-09-858-332G-28 (1-6) x AC149204 (1-75741)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       8080 CATAATCATAATCATAAT 8097
```

```
RESULT 494
AC037424/c
LOCUS       AC037424               75948 bp    DNA     linear   PLN 19-JAN-2001
DEFINITION  Arabidopsis thaliana chromosome 1 BAC F19K6 genomic sequence,
            complete sequence.
ACCESSION   AC037424
VERSION     AC037424.10  GI:12323115
KEYWORDS    HTG.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 75948)
  AUTHORS   Lin,X., Kaul,S., Town,C.D., Benito,M.-I., Creasy,T.H., Haas,B.J.,
            Wu,D., Maiti,R., Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R.,
            Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.
  TITLE     Arabidopsis thaliana chromosome 1 BAC F19K6 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 75948)
  AUTHORS   Lin,X. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
REFERENCE   3  (bases 1 to 75948)
  AUTHORS   Town,C.D. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
COMMENT     On Jan 19, 2001 this sequence version replaced gi:12280784.
            Address all correspondence to:at@tigr.org

            BAC clone F19K6 is from Arabidopsis thaliana chromosome 1
            The orientation of the sequence is from SP6 to T7 end of the BAC
            clone.
            Genes were identified by a combination of several methods: Gene
            prediction programs including Genscan+ (Chris Burge,
            http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHMM (Mark Borodovsky,
            http://genemark.biology.gatech.edu/GeneMark/), GlimmerA (a variant
            of GlimmerM, see Mihaela Pertea,
            http://www.tigr.org/softlab/glimmerm_htm/glimmerm.html, and
            GeneSplicer (Mihaela Pertea and Steven Salzberg, contact
            mpertea@tigr.org), searches of the complete sequence against a
            peptide database and the plant EST database at TIGR
            (http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to
            indicate the level of evidence for their annotation. Genes with
            similarity to other proteins are named after the database hits.
            Genes without significant peptide similarity but with EST
            similarity are named as  unknown  proteins. Genes without protein
            or EST similarity, that are predicted by more than two gene
            prediction programs over most of their length are annotated as
            hypothetical  proteins. Genes encoding tRNAs are predicted by
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
            Simple repeats are identified by repeatmasker (Arian Smit,
            http://ftp.genome.washington.edu/RM/RepeatMasker.html).
FEATURES             Location/Qualifiers
     source          1. .75948
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /chromosome="I"
                     /map="m213"
                     /clone="F19K6"
                     /ecotype="Columbia"
     gene            1. .1723
                     /gene="F19K6.1"
     mRNA            join(<1. .97,223. .405,517. .723,811. .958,1057. .1187,
                     1271. .1360,1454. .1534,1628. .>1723)
                     /gene="F19K6.1"
     CDS             join(<1. .97,223. .405,517. .723,811. .958,1057. .1187,
                     1271. .1360,1454. .1534,1628. .1723)
                     /gene="F19K6.1"
                     /codon_start=2
                     /product="unknown protein, 5' partial; 1-1723"
                     /protein_id="AAG51537.1"
                     /db_xref="GI:12323117"
                     /translation="DMPKEDSLSAHSESAIGGAAARYNSALDMTLKALLNIKVLKEDL
                     KNNTQRFQQVPSALQHFFNAFVSESIKTEGVYSCILSDLLSSLEEVISMSSDAAKVVV
                     AILEFWCCWKNPEGESLVTRLFTLAENERMSCRKCRWITNYPEKSSYAIVMPADSIRQ
                     LKCAFESMKFVDILKVIRMEYQKSCNIKTGGCGESNFVHHFISRCPPVFTIVLEWEKS
                     ETEKEISETTKALDLELDISRLYVGLEPNTNYRLVSMVGCGEEEEHFCLAYENNRWVN
                     LGRESLAGEDVSNWKSEVIFFGDFKVNLRQESLSGEDDGNWKSVVSFFGERKIRPEIL
                     FYEAVGSMA"
gene                 complement(1815. .4406)
                     /gene="F19K6.13"
                     /note="similar to Pfam profile: PF00534 Glycosyl
                     transferases group 1"
mRNA                 complement(join(<1815. .2337,2604. .3364,3455. .>4406))
                     /gene="F19K6.13"
CDS                  complement(join(2038. .2337,2604. .3364,3455. .4406))
                     /gene="F19K6.13"
                     /codon_start=1
                     /product="glycosyl transferase, putative; 4406-2038"
                     /protein_id="AAG51540.1"
                     /db_xref="GI:12323120"
                     /translation="MEEIRLSPLRQTSVKSSLSGRSTPRGTPRVYSGRTPRRGHGGGG
                     AFQWFRSSRLVYWLLLITLWTYLGFYVQSRWAHDNESKVEFLRFGGRPRKDELYMEKI
                     KGLDVVANENSEALVNITGKDDAGSNKRTDVSLIKKDDGVSRRSLSSKQKTRKTVRTS
                     RSKIRVKQKVIKEVLETKDLDDEQDPQLPLTNATYGKLLGPFGSLEDKVLEWSPHRRS
                     GTCDRKSDFKRLVWSRRFVLLFHELSMTGAPISMMELASELLSCGATVSAVVLSRRGG
                     LMQELSRRRIKVVEDKGELSFKTAMKADLIIAGSAVCTSWIDQYMNHHPAGGSQIAWW
                     IMENRREYFDRAKPVLDRVKMLIFLSESQSRQWLTWCEEEHIKLRSQPVIVPLSVNDE
                     LAFVAGIPSSLNTPTLSPEKMRVKRQILRESVRTELGITDSDMLVMSLSSINPTKGQL
                     LLLESIALALSERGQESQRNHKGIIRKEKVSLSSKHRLRGSSRQMKSVSLTLDNGLRR
                     EKQELKVLLGSVGSKSNKVGYVKEMLSFLSNSGNLSKSVMWTPATTRVASLYSAADVY
                     VTNSQGVGETFGRVTIEAMAYGLAVVGTDAGGTKEMVQHNMTGLLHSMGRSGNKELAH
                     NLLYLLRNPDERLRLGSEGRKMVEKMYMKQHMYKRFVDVLVKCMRP"
gene                 complement(7506. .12311)
                     /gene="F19K6.14"
                     /note="contains Pfam profile: PF00658 Poly-adenylate
                     binding protein, unique domain."
mRNA                 complement(join(<7506. .8416,8508. .8662,8747. .8788,
                     8881. .8928,9012. .9071,9155. .9202,9286. .9354,9431. .9478,
                     9572. .9634,9716. .9904,10002. .10049,10147. .10215,
                     10314. .10361,10431. .10499,10624. .10671,10758. .10826,
                     10927. .10974,11063. .11128,11228. .11275,11359. .11433,
                     11574. .11643,11784. .11869,11953. .11994,12093. .12144,
                     12222. .>12311))
                     /gene="F19K6.14"
CDS                  complement(join(7712. .8416,8508. .8662,8747. .8788,
                     8881. .8928,9012. .9071,9155. .9202,9286. .9354,9431. .9478,
                     9572. .9634,9716. .9904,10002. .10049,10147. .10215,
                     10314. .10361,10431. .10499,10624. .10671,10758. .10826,
                     10927. .10974,11063. .11128,11228. .11275,11359. .11433,
                     11574. .11643,11784. .11869,11953. .11994,12093. .12144,
                     12222. .12311))
                     /gene="F19K6.14"
                     /codon_start=1
                     /product="myosin-like protein; 12311-7712"
                     /protein_id="AAG51543.1"
                     /db_xref="GI:12323123"
                     /translation="MEIYTMKTNFLVLALSLCILLSSFHEVSCQDDGSGLSNLDLIER
                     DYQDSVNALQGKDDEDQSAKIQSENQNNTTVTDKNTISLSLSDESEVGSVSDESVGRS
                     SLLDQIKLSFEAHHNSINQAGSDGVKAESKDDDEELSAHRQKMLEEIEHEFEAASDSL
                     KQLKTDDVNEGNDEEHSAKRQSLLEEIEREFEAATKELEQLKVNDFTGDKDDEEHSAK
                     RKSMLEAIEREFEAAMEGIEALKVSDSTGSGDDEEQSAKRLSMLEEIEREFEAASKGL
                     EQLRASDSTADNNEEEHAAKGQSLLEEIEREFEAATESLKQLQVDDSTEDKEHCKALF
                     FLLSAILSLWLSESGFECIVVTAAKRQSLLEEIEREFEAATKDLKQLNDFTEGSADDE
                     QSAKRNKMLEDIEREFEAATIGLEQLKANDFSEGNNNEEQSAKRKSMLEEIEREFEAA
                     IGGLKQIKVDDSRNLEEESAKRKIILEEMEREFEEAHSGINAKADKEESAKKQSGSAI
                     PEVLGLGQSGGCSCSKQDEDSSIVIPTKYSIEDILSEESAVQGTETSSLTASLTQLVE
                     NHRKEKESLLGHRVLTSPSIASSTSESSATSETVETLRAKLNELRGLTARELVTRKDF
                     GQILITAASFEELSSAPISYISRLAKYRNVIKEGLEASERVHIAQVRAKMLKEVATEK
                     QTAVDTHFATAKKLAQEGDALFVKIFAIKKLLAKLEAEKESVDGKFKETVKELSHLLA
                     DASEAYEEYHGAVRKAKDEQAAEEFAKEATQSAEIIWVKFLSSL"
```

```
     gene            complement(14967. .17884)
                     /gene="F19K6.15"
                     /note="identical to GI:6651430 from [Arabidopsis
                     thaliana]"
     mRNA            complement(join(<14967. .15347,15431. .15542,15637. .15739,
                     15870. .15904,16005. .16219,16303. .16412,16496. .16848,
                     16941. .17018,17169. .17244,17341. .17396,17478. .17547,
                     17650. .>17884))
                     /gene="F19K6.15"
     CDS             complement(join(15143. .15347,15431. .15542,15637. .15739,
                     15870. .15904,16005. .16219,16303. .16412,16496. .16848,
                     16941. .17018,17169. .17244,17341. .17396,17478. .17547,
                     17650. .17823))
                     /gene="F19K6.15"
                     /codon_start=1
                     /product="beta-glucosidase, putative; 17823-15143"
                     /protein_id="AAG51546.1"
                     /db_xref="GI:12323126"
                     /translation="MVRFEKVHLVLGLALVLTLVGAPTKAQGPVCGAGLPDKFSRLNF
                     PEGFIWGTATAAFQVEGAVNEGCRGPSMWDTFTKKFPHRCENHNADVAVDFYHRYKED
                     IQLMKDLNTDAFRLSIAWPRIFPHGRMSKGISKVGVQFYHDLIDELLKNNIIPLVTVF
                     HWDTPQDLEDEYGGFLSGRIVQDFTEYANFTFHEYGHKVKHWITFNEPWVFSRAGYDN
                     GKKAPGRCSPYIPGYGQHCQDGRSGYEAYQVSHNLLLSHAYAVDAFRNCKQCAGGKIG
                     IAHSPAWFEPQDLEHVGGSIERVLDFILGWHLAPTTYGDYPQSMKDRVGHRLPKFTEA
                     EKKLLKGSTDYVGMNYYTSVFAKEISPDPKSPSWTTDSLVDWDSKSVDGYKIGSKPFN
                     GKLDVYSKGLRYLLKYIKDNYGDPEVIIAENGYGEDLGEKHNDVNFGTQDHNRKYYIQ
                     RHLLSMHDAICKDKVNVTGYFVWSLMDNFEWQDGYKARFGLYYIDFQNNLTRHQKVSG
                     KWYSEFLKPQFPTSKLREEL"
     gene            20839. .21105
                     /gene="F19K6.2"
                     /note="predicted by genemark.hmm"
     mRNA            join(<20839. .21024,21076. .>21105)
                     /gene="F19K6.2"
     CDS             join(20839. .21024,21076. .21105)
                     /gene="F19K6.2"
                     /codon_start=1
                     /product="hypothetical protein; 20839-21105"
                     /protein_id="AAG51548.1"
                     /db_xref="GI:12323128"
                     /translation="MVDIVCWAVKNGSAGRNLLLIANASSEDDQFWPFLQGLRYRGFK
                     LFATIPDRTLPDDECLTAGIEDTGSDD"
     gene            complement(21772. .24076)
                     /gene="F19K6.4"
     mRNA            complement(join(<21772. .23135,23968. .24076))
                     /gene="F19K6.4"
     CDS             complement(21772. .23094)
                     /gene="F19K6.4"
                     /codon_start=1
                     /product="unknown protein; 23094-21772"
                     /protein_id="AAG51549.1"
                     /db_xref="GI:12323129"
                     /translation="MGDSENVQQPSKKRGALKQLSRDNPGLDDDDDSAELESGTFKKA
                     SDEVLASRRIVRVKRKEPSAAPVAASNPFAGIQLVPTTAPASTPVGTNAPLAESKLAP
                     AEAVVEDNQKASDIEEGDEVDSKKVDVKDAVGEETEKTKDKDDNHCGKSADVQVAATE
                     VAQMVSCDTNVCNNAVEGTDQTDFPLEKDSGGDQAEKKEKEGNGIEEADKNGDNGAFS
                     SFQQHSSNKNAFTGLASTEASGSSFSFGLVSQDGSTGTGSLFGFGLPSSNSSSIFGAT
                     GSSIIKKSEGSGFPPKQEVSTETGEENEKVAFSADSIMFEYLDGGWKERGKGELKVNV

Alignment Scores:
Pred. No.:              1.51e+04        Length:         75948
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x AC037424 (1-75948)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      47883 CACAACCACAACCACAAC 47866

RESULT 495
```

```
AC101697
LOCUS       AC101697               76240 bp    DNA     linear   HTG 23-NOV-2001
DEFINITION  Mus musculus clone RP23-240E1, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC101697
VERSION     AC101697.1  GI:17060472
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 76240)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-240E1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 76240)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L17239
                Center clone name: 240_E_1
            --------------.
            * NOTE: This record contains 92 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      703: contig of 703 bp in length
            *      704      803: gap of 100 bp
            *      804     1542: contig of 739 bp in length
            *     1543     1642: gap of 100 bp
            *     1643     2377: contig of 735 bp in length
            *     2378     2477: gap of 100 bp
            *     2478     3217: contig of 740 bp in length
            *     3218     3317: gap of 100 bp
            *     3318     4020: contig of 703 bp in length
```

```
*     4021     4120: gap of 100 bp
*     4121     4837: contig of 717 bp in length
*     4838     4937: gap of 100 bp
*     4938     5688: contig of 751 bp in length
*     5689     5788: gap of 100 bp
*     5789     6503: contig of 715 bp in length
*     6504     6603: gap of 100 bp
*     6604     7348: contig of 745 bp in length
*     7349     7448: gap of 100 bp
*     7449     8147: contig of 699 bp in length
*     8148     8247: gap of 100 bp
*     8248     8975: contig of 728 bp in length
*     8976     9075: gap of 100 bp
*     9076     9816: contig of 741 bp in length
*     9817     9916: gap of 100 bp
*     9917    10621: contig of 705 bp in length
*    10622    10721: gap of 100 bp
*    10722    11422: contig of 701 bp in length
*    11423    11522: gap of 100 bp
*    11523    12259: contig of 737 bp in length
*    12260    12359: gap of 100 bp
*    12360    13101: contig of 742 bp in length
*    13102    13201: gap of 100 bp
*    13202    13942: contig of 741 bp in length
*    13943    14042: gap of 100 bp
*    14043    14775: contig of 733 bp in length
*    14776    14875: gap of 100 bp
*    14876    15608: contig of 733 bp in length
*    15609    15708: gap of 100 bp
*    15709    16446: contig of 738 bp in length
*    16447    16546: gap of 100 bp
*    16547    17270: contig of 724 bp in length
*    17271    17370: gap of 100 bp
*    17371    18097: contig of 727 bp in length
*    18098    18197: gap of 100 bp
*    18198    18933: contig of 736 bp in length
*    18934    19033: gap of 100 bp
*    19034    19767: contig of 734 bp in length
*    19768    19867: gap of 100 bp
*    19868    20536: contig of 669 bp in length
*    20537    20636: gap of 100 bp
*    20637    21376: contig of 740 bp in length
*    21377    21476: gap of 100 bp
*    21477    22171: contig of 695 bp in length
*    22172    22271: gap of 100 bp
*    22272    23013: contig of 742 bp in length
*    23014    23113: gap of 100 bp
*    23114    23848: contig of 735 bp in length
*    23849    23948: gap of 100 bp
*    23949    24672: contig of 724 bp in length
*    24673    24772: gap of 100 bp
*    24773    25511: contig of 739 bp in length
*    25512    25611: gap of 100 bp
*    25612    26319: contig of 708 bp in length
*    26320    26419: gap of 100 bp
*    26420    27161: contig of 742 bp in length
*    27162    27261: gap of 100 bp
*    27262    28004: contig of 743 bp in length
*    28005    28104: gap of 100 bp
*    28105    28841: contig of 737 bp in length
*    28842    28941: gap of 100 bp
*    28942    29690: contig of 749 bp in length
*    29691    29790: gap of 100 bp
*    29791    30450: contig of 660 bp in length
*    30451    30550: gap of 100 bp
*    30551    31289: contig of 739 bp in length
*    31290    31389: gap of 100 bp
*    31390    32135: contig of 746 bp in length
*    32136    32235: gap of 100 bp
*    32236    32963: contig of 728 bp in length
*    32964    33063: gap of 100 bp
*    33064    33799: contig of 736 bp in length
*    33800    33899: gap of 100 bp
*    33900    34641: contig of 742 bp in length
*    34642    34741: gap of 100 bp
*    34742    35477: contig of 736 bp in length
*    35478    35577: gap of 100 bp
*    35578    36323: contig of 746 bp in length
*    36324    36423: gap of 100 bp
*    36424    37165: contig of 742 bp in length
*    37166    37265: gap of 100 bp
*    37266    37997: contig of 732 bp in length
*    37998    38097: gap of 100 bp
*    38098    38836: contig of 739 bp in length
*    38837    38936: gap of 100 bp
*    38937    39682: contig of 746 bp in length
*    39683    39782: gap of 100 bp
*    39783    40500: contig of 718 bp in length
*    40501    40600: gap of 100 bp
*    40601    41329: contig of 729 bp in length
*    41330    41429: gap of 100 bp
*    41430    42172: contig of 743 bp in length
*    42173    42272: gap of 100 bp
*    42273    43007: contig of 735 bp in length
*    43008    43107: gap of 100 bp
*    43108    43828: contig of 721 bp in length
*    43829    43928: gap of 100 bp
*    43929    44674: contig of 746 bp in length
*    44675    44774: gap of 100 bp
*    44775    45466: contig of 692 bp in length
*    45467    45566: gap of 100 bp
*    45567    46311: contig of 745 bp in length
*    46312    46411: gap of 100 bp
*    46412    47162: contig of 751 bp in length
*    47163    47262: gap of 100 bp
*    47263    47998: contig of 736 bp in length
*    47999    48098: gap of 100 bp
*    48099    48840: contig of 742 bp in length
*    48841    48940: gap of 100 bp
*    48941    49650: contig of 710 bp in length
*    49651    49750: gap of 100 bp
*    49751    50492: contig of 742 bp in length
*    50493    50592: gap of 100 bp
*    50593    51313: contig of 721 bp in length
*    51314    51413: gap of 100 bp
*    51414    52130: contig of 717 bp in length
*    52131    52230: gap of 100 bp
*    52231    52954: contig of 724 bp in length
*    52955    53054: gap of 100 bp
*    53055    53776: contig of 722 bp in length
*    53777    53876: gap of 100 bp
*    53877    54601: contig of 725 bp in length
*    54602    54701: gap of 100 bp
*    54702    55411: contig of 710 bp in length
*    55412    55511: gap of 100 bp
*    55512    56243: contig of 732 bp in length
*    56244    56343: gap of 100 bp
*    56344    57084: contig of 741 bp in length

Alignment Scores:
Pred. No.:                1.52e+04        Length:          76240
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       14              Gaps:            0

US-09-858-332G-28 (1-6) x AC101697 (1-76240)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      58362 CATAACCATAACCATAAC 58379

RESULT 496
AC023249/c
LOCUS       AC023249               76336 bp    DNA     linear   HTG 13-JUL-2000
```

```
DEFINITION  Homo sapiens chromosome 6 clone RP11-366G3 map 6, LOW-PASS SEQUENCE
            SAMPLING.
ACCESSION   AC023249
VERSION     AC023249.2  GI:9156338
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 76336)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 6, clone RP11-366G3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 76336)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 13, 2000 this sequence version replaced gi:6958054.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L6053
                Center clone name: 366_G_3
            --------------.
            * NOTE: This record contains 80 individual   *
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      831: contig of 831 bp in length
            *      832      931: gap of 100 bp
            *      932     1803: contig of 872 bp in length
            *     1804     1903: gap of 100 bp
            *     1904     2772: contig of 869 bp in length
            *     2773     2872: gap of 100 bp
            *     2873     3738: contig of 866 bp in length
            *     3739     3838: gap of 100 bp
            *     3839     4686: contig of 848 bp in length
            *     4687     4786: gap of 100 bp
            *     4787     5625: contig of 839 bp in length
            *     5626     5725: gap of 100 bp
            *     5726     6574: contig of 849 bp in length
            *     6575     6674: gap of 100 bp
            *     6675     7539: contig of 865 bp in length
            *     7540     7639: gap of 100 bp
            *     7640     8489: contig of 850 bp in length
            *     8490     8589: gap of 100 bp
            *     8590     9436: contig of 847 bp in length
            *     9437     9536: gap of 100 bp
            *     9537    10367: contig of 831 bp in length
            *    10368    10467: gap of 100 bp
            *    10468    11292: contig of 825 bp in length
            *    11293    11392: gap of 100 bp
            *    11393    12268: contig of 876 bp in length
            *    12269    12368: gap of 100 bp
            *    12369    13234: contig of 866 bp in length
            *    13235    13334: gap of 100 bp
            *    13335    14202: contig of 868 bp in length
            *    14203    14302: gap of 100 bp
            *    14303    15142: contig of 840 bp in length
            *    15143    15242: gap of 100 bp
            *    15243    16072: contig of 830 bp in length
            *    16073    16172: gap of 100 bp
            *    16173    17042: contig of 870 bp in length
            *    17043    17142: gap of 100 bp
            *    17143    17972: contig of 830 bp in length
            *    17973    18072: gap of 100 bp
            *    18073    18948: contig of 876 bp in length
            *    18949    19048: gap of 100 bp
            *    19049    19925: contig of 877 bp in length
            *    19926    20025: gap of 100 bp
            *    20026    20847: contig of 822 bp in length
            *    20848    20947: gap of 100 bp
            *    20948    21904: contig of 957 bp in length
            *    21905    22004: gap of 100 bp
            *    22005    22862: contig of 858 bp in length
            *    22863    22962: gap of 100 bp
            *    22963    23838: contig of 876 bp in length
            *    23839    23938: gap of 100 bp
            *    23939    24767: contig of 829 bp in length
            *    24768    24867: gap of 100 bp
            *    24868    25696: contig of 829 bp in length
            *    25697    25796: gap of 100 bp
            *    25797    26627: contig of 831 bp in length
            *    26628    26727: gap of 100 bp
            *    26728    27575: contig of 848 bp in length
            *    27576    27675: gap of 100 bp
            *    27676    28537: contig of 862 bp in length
            *    28538    28637: gap of 100 bp
            *    28638    29493: contig of 856 bp in length
            *    29494    29593: gap of 100 bp
            *    29594    30414: contig of 821 bp in length
            *    30415    30514: gap of 100 bp
            *    30515    31354: contig of 840 bp in length
            *    31355    31454: gap of 100 bp
            *    31455    32296: contig of 842 bp in length
            *    32297    32396: gap of 100 bp
            *    32397    33270: contig of 874 bp in length
            *    33271    33370: gap of 100 bp
            *    33371    34255: contig of 885 bp in length
            *    34256    34355: gap of 100 bp
            *    34356    35230: contig of 875 bp in length
            *    35231    35330: gap of 100 bp
            *    35331    36168: contig of 838 bp in length
            *    36169    36268: gap of 100 bp
            *    36269    37099: contig of 831 bp in length
            *    37100    37199: gap of 100 bp
            *    37200    38025: contig of 826 bp in length
            *    38026    38125: gap of 100 bp
            *    38126    38998: contig of 873 bp in length
            *    38999    39098: gap of 100 bp
            *    39099    39957: contig of 859 bp in length
            *    39958    40057: gap of 100 bp
            *    40058    40895: contig of 838 bp in length
            *    40896    40995: gap of 100 bp
            *    40996    41825: contig of 830 bp in length
```

```
*    41826    41925: gap of 100 bp
*    41926    42836: contig of 911 bp in length
*    42837    42936: gap of 100 bp
*    42937    43765: contig of 829 bp in length
*    43766    43865: gap of 100 bp
*    43866    44732: contig of 867 bp in length
*    44733    44832: gap of 100 bp
*    44833    45685: contig of 853 bp in length
*    45686    45785: gap of 100 bp
*    45786    46630: contig of 845 bp in length
*    46631    46730: gap of 100 bp
*    46731    47557: contig of 827 bp in length
*    47558    47657: gap of 100 bp
*    47658    48503: contig of 846 bp in length
*    48504    48603: gap of 100 bp
*    48604    49449: contig of 846 bp in length
*    49450    49549: gap of 100 bp
*    49550    50392: contig of 843 bp in length
*    50393    50492: gap of 100 bp
*    50493    51335: contig of 843 bp in length
*    51336    51435: gap of 100 bp
*    51436    52285: contig of 850 bp in length
*    52286    52385: gap of 100 bp
*    52386    53244: contig of 859 bp in length
*    53245    53344: gap of 100 bp
*    53345    54142: contig of 798 bp in length
*    54143    54242: gap of 100 bp
*    54243    55113: contig of 871 bp in length
*    55114    55213: gap of 100 bp
*    55214    56045: contig of 832 bp in length
*    56046    56145: gap of 100 bp
*    56146    56985: contig of 840 bp in length
*    56986    57085: gap of 100 bp
*    57086    57929: contig of 844 bp in length
*    57930    58029: gap of 100 bp
*    58030    58875: contig of 846 bp in length
*    58876    58975: gap of 100 bp
*    58976    59808: contig of 833 bp in length
*    59809    59908: gap of 100 bp
*    59909    60744: contig of 836 bp in length
*    60745    60844: gap of 100 bp
*    60845    61679: contig of 835 bp in length
*    61680    61779: gap of 100 bp
*    61780    62645: contig of 866 bp in length
*    62646    62745: gap of 100 bp
*    62746    63612: contig of 867 bp in length
*    63613    63712: gap of 100 bp
*    63713    64553: contig of 841 bp in length
*    64554    64653: gap of 100 bp
*    64654    65499: contig of 846 bp in length
*    65500    65599: gap of 100 bp
*    65600    66458: contig of 859 bp in length
*    66459    66558: gap of 100 bp

Alignment Scores:
Pred. No.:                1.52e+04        Length:          76336
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       14              Gaps:            0

US-09-858-332G-28 (1-6) x AC023249 (1-76336)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      75224 CACAACCACAACCACAAC 75207

RESULT 497
AC147603
LOCUS       AC147603               78404 bp    DNA     linear   HTG 17-AUG-2004
DEFINITION  Zea mays clone ZMMBBc0307D13, *** SEQUENCING IN PROGRESS ***, 2
            ordered pieces.
ACCESSION   AC147603
VERSION     AC147603.5  GI:51315592
KEYWORDS    HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 78404)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Butler,E., Wing,R., Bharti,A.K.
            and Messing,J.
  TITLE     Zea mays, clone ZMMBBc0307D13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 78404)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-DEC-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 78404)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Butler,E., Wing,R., Bharti,A.K.,
            Messing,J., Abouelleil,A., Allen,N., Anderson,M., Anderson,S.,
            Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., Corum,B., DeArellano,K., Diaz,J.S., Dodge,S.,
            Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P.,
            FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L.,
            Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., Liu,X., Lui,A., Mabbitt,R., MacLean,C., Macdonald,P.,
            Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Nguyen,T., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-AUG-2004) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Aug 17, 2004 this sequence version replaced gi:47155670.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@broad.mit.edu
            --------------
                Bharti, AK and Messing, J: The Plant Genome Initiative at
```

Rutgers, Waksman Institute, Rutgers, The State University of New Jersey, 190 Frelinghuysen Road, Piscataway, NJ 08854, USA (http://pgir.rutgers.edu)
Butler, E and Wing, R: Arizona Genomics Institute, Biological Sciences West, 448A, P.O. Box 210088, University of Arizona, Tucson, AZ 85721, USA (http://www.genome.arizona.edu)
---------- Project Information
Center project name: L30037
Center clone name: 307_D_13
---------- Consensus Information
This consensus is derived from a shotgun assembly that has been manually curated. It is the best representation of the BAC that we can generate without further laboratory work. The draft assembly has been edited, and if possible, ends identified by vector as well as by BAC end sequences, and contigs ordered and oriented. Bases that are not Ns are either above Q20 or manually edited. This assembly was performed with Arachne (Genome Res. 2002 12: 177-189; Genome Res. 2003 13: 91-96). All trace files for this project are available at the NCBI trace repository (http://www.ncbi.nlm.nih.gov/Traces/trace.fcgi?). An exact list of reads used in this assembly are available at http://www.broad.mit.edu/annotation/plants/maize/randomclones.html.
----------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 17470: contig of 17470 bp in length
* 17471 17570: gap of unknown length
* 17571 78404: contig of 60834 bp in length.

FEATURES Location/Qualifiers
source 1. .78404
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="B73"
/db_xref="taxon:4577"
/clone="ZMMBBc0307D13"
/clone_lib="CHORI-201 Maize B73 BAC MboI Library"
gap 17471. .17570
/estimated_length=unknown

ORIGIN

Alignment Scores:

Pred. No.:	1.56e+04	Length:	78404
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	14	Gaps:	0

US-09-858-332G-28 (1-6) x AC147603 (1-78404)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      70203 CACAACCACAACCACAAC 70220
```

RESULT 498
AB024035
LOCUS AB024035 78423 bp DNA linear PLN 14-FEB-2004
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone:MHM17.
ACCESSION AB024035 BA000015
VERSION AB024035.1 GI:4519194
KEYWORDS .
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1
AUTHORS Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E., Kotani,H. and Tabata,S.
TITLE Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence features of the regions of 3,076,755 bp covered by sixty P1 and TAC clones
JOURNAL DNA Res. 7 (1), 31-63 (2000)
PUBMED 10718197
REFERENCE 2 (bases 1 to 78423)
AUTHORS Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (24-FEB-1999) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp, Tel:81-438-52-3935, Fax:81-438-52-3934)
COMMENT Address for correspondence: kaos@kazusa.or.jp
For the latest information on annotation of this clone, please see http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MHM17
Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'.
The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, http://compbio.ornl.gov/Grail-1.3/), GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html), NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and SplicePredictor (Volker Brendel, Stanford University, http://gremlin1.zool.iastate.edu/cgi-bin/sp.cgi).
Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, http://genome.wustl.edu/eddy/tRNAscan-SE/).
This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions. The 5' clone is MPI10 and the 3' clone is MUL3.

FEATURES Location/Qualifiers
source 1. .78423
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/chromosome="5"
/clone="MHM17"
/clone_lib="Mitsui P1"
/ecotype="Columbia"
exon complement(1. .30)
/note="CDS is reported in Acc# AB020747 contains similarity to protein kinase gene_id:MPI10.5"
/number=1
/evidence=not_experimental
CDS complement(join(4221. .4349,4616. .4761,4851. .4944, 5022. .5307,5397. .5821,6091. .6208,6305. .6432,6529. .6638, 6709. .6982,7070. .7141))
/note="unnamed protein product; dbj|BAA91947.1 gene_id:MHM17.1 similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAA97020.1"
/db_xref="GI:8777430"
/translation="MAPRILLCGDPLGRLNQLFKRVQSVSKSAGPFDALICVGQFFPD SPEILDEFLDYVEGRAQVPIPTYFTGDYGVVAPKILSTTSKKAENQGFKMDGLEVCHN LFWLRGSGKFSLHGLSVAYLSGRQSSESQFGKYSQDDVDALRALAEEPGGCITNEWPA GVTNRAAVSDIPVGISDSSCSDSTVSELVMEVKPRYHIAGSMGVFYAREPYLNAESTH VTRFIGLAQVGNKNKQKFLHALSPTPTSTMSPAELSAKPPKTTLWPYNLQDGAAESKK RPNDSDSDSQYWRYDVPKRQKSGSQGEKLCFKFVCSGSCPRGEDCHFQHNAEAREQCR RGVCLDLIIKGKCEKGPECSYKHEFQDESSIQRKPRSENANRSKECWFCLSSPSVESH LIVSVGESFYCALPKGSLVEDHILIIPIEHLPNTLVLSPEVESELSRYQNGLRNCYKS QGNDAVFFELVSKRVSHANLQVVPVPSSRARLLPNIFSLAAEKLGFKLVTKKFNDSTD GRKYLQKEYNAALGLFYVELPDGTVLSHTLEENEVFPAQFGREVLAGLLKIPDRADWR NCKISQEEEAKLAEDFKKQFQEFDPCQ"
CDS join(7467. .7643,7741. .8080,8177. .8230,8376. .8479)
/note="unnamed protein product; gene_id:MHM17.2

```
                pir||C71422
                similar to unknown protein"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="BAA97021.1"
                /db_xref="GI:8777431"
                /translation="MADSVLKEVNCGRPEKIPKLDKACEGSKSSWKHLKLGNVEDDEY
                LRQYCLFHYEFHKSEGFTVDWEKYDYMFHIRPLENSPPISDIRTNADVIRDVTLFAIE
                KHNEAHGSKLVFVEHVSANFKFANGLLCWLTFWATDMASSAPTSQIYQVELWRRGKQF
                EIPIFRVKPKDEEMDDVEVKPPSPMPYDDYDKPPVVFVRAAPEDGVPFVFDRTGALYD
                LYRSGL"
CDS             join(9471. .9653,9736. .10236)
                /note="unnamed protein product; gene_id:MHM17.3
                pir||C71422
                similar to unknown protein"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="BAA97022.1"
                /db_xref="GI:8777432"
                /translation="MADSLLMEVNYRRPEKIPKLDEDCEGSESSSKQRVVKWGTAEDD
                EYLRQYLLFHYQFQKTQGFSIKWEQFDYNFRSRSMDKSPNEISERKSNAVLIREMTLT
                AIDKHNEAHGTKLVFVEHVEANYQLTRGLTCWLTFWATDMASSPPQSKIYQAHVWRRG
                QNFHTFIFRLKPTDEEIEAVEVQPPSPMLYELDKPPIVFSRPGPEDEALPGVPFVFNR
                TGAGLDPDW"
CDS             complement(join(12420. .13157,13254. .13394,13492. .13542,
                13628. .13759,13862. .13927,14005. .14747,14880. .15036))
                /note="unnamed protein product; gene_id:MHM17.4
                unknown protein"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="BAA97023.1"
                /db_xref="GI:8777433"
                /translation="MESSVAPFPHRRTHLPNRNYRSLLYHFCSSFDREPQISLATPAV
                LQELVLRTEIAQESPGETCEPPENLSITESKLNGVSGDSSGEKVETISQEKSLMLGDI
                CDGIDLQDASVVSRHTDFFDSFELMINETQDSVPESCVNLFEALDVNDYDIVQNVLEK
                PNIATQVQVDPVESEKKAEEVPKSVESNEVISSGVLEACNGTVQREMELEKPVDNSPV
                LVDSVSRIVGGDDVEEGEISGDDNDDMLVEDDETVERHEEYQVSQDGTGNSHLTSHKS
                FGVEVMNVDNQAKKIDQTFSNEAKMDPGTSIKKRSAPSKDAKARKRAKARIKRAQERI
                ALGVKKLKLKPVAPKPKPIKYCRHYLKGRCHEGDKCKFSHDTIPETKCSPCCYFATQS
                CMKGDDCPFDHDLSKYPCNNFITKGFCYRGDSCLFSHKGTPQSASDTPSANVTVSSTK
                ITAASFSPQKTKKQSVRDAIAKLPAIQARVSSSVAFLKPSSHSNQRNSSDASSSKINE
                HVTPPQVPPLRKPSVAPKGMSFLSLDKTSQEDTVKASSASKPNTDNSDSQTLKQSQQG
                SFLPLGPPKGISFLSFASEEQKTLNREPQKPASSKNLKTTPSSHIQSSLLSAMKLAAE
                FESAKVERGNNDPTEAVNKSNVTVDTAVTRNSGNISSKILEFLSSFSHGKN"
CDS             join(15350. .15482,15570. .15615,16443. .16671)
                /note="gene_id:MHM17.5"
                /codon_start=1
                /evidence=not_experimental
                /product="30S ribosomal protein S16"
                /protein_id="BAA97024.1"
                /db_xref="GI:8777434"
                /translation="MVVRIRLSRFGCKNRPFFRVMAADSRSPRDGKHLEVLGYFNPLP
                GQDGGKRMGLKFDRIKYWLSVGAQPSDPVQRLLFRSGLLPPPPMVAMGRKGGARDTRP
                VDPMTGRYVDAENKTVNANDNQPKEEDTEAKSA"
CDS             join(17089. .17134,17217. .17260,17389. .17484,17694. .17786,
                17919. .18006,18095. .18156,18245. .18429,18562. .18739,
                18827. .18870,19017. .19218,19684. .19770)
                /note="gene_id:MHM17.6"
                /codon_start=1
                /evidence=not_experimental
                /product="nucleosome assembly protein"
                /protein_id="BAA97025.1"
                /db_xref="GI:8777435"
                /translation="MSNDKDSFNVSDLTSALKDEDRAGLVNALKNKLQNLAGQHSDVL
                ENLTPKIRRRVEVLREIQGKHDEIETKFREERAALEAKYQKLYQPLYNKRYEIVNGAT
                EVEGAPEDAKMDQGDEKTAEEKGVPSFWLTALKNNDVISEEITERDEGALIYLKDIKW
                CKIEEPKGFKLEFFFDQNPYFKNTLLTKAYHMIDEDEPLLEKAIGTEIDWYPGKCLTQ
                KILKKKPKKGAKNAKPITKTEDCESFFNFFNPPQVPDDDEDIDEERAEELQNLMEQDY
                DIGSTIREKIIPHAVSWFTGEAIEGEEFEIDNDDEDDIDEDEDEDEEDEDEEEDDE
                DEEEEVSKTKKKFEIILQKGGRPQVTDDQQGERPPECKQQ"
CDS             complement(join(23174. .23239,23666. .24049,24174. .24863,
                25283. .25388,25491. .25645))
                /note="unnamed protein product; gb|AAC78547.1
                gene_id:MHM17.7
                similar to unknown protein"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="BAA97026.1"
                /db_xref="GI:8777436"
                /translation="MMHLILSCSYLISMDGYYNEASEEPSSSSSSGSLARSLFHEYRQ
                SVIPLQNGHVPSMAFMNNLPYVEIRPQESQRLAFNDTQRLFYQMKIEASLREWFPEDF
                NRKSSPANSDYLRPPHYPSSSSSSLSPNNISEYSSLLFPLIPKPSTTTEAVNVPVLPP
                LAPINMIHPQHQEPLFRNRQREEEAMTQAILAVLTGPSSPPSTSSSPQRKGRATAFKR
                YYSMISDRGRAPLPSVRKQSMMTRAMSFYNRLNINQRERFTRENATTHGEGSGGSGGG
                GRYTSGPSATQLQHMISERKRREKLNESFQALRSLLPPGTKKDKASVLSIAREQLSSL
                QGEISKLLERNREVEAKLAGEREIENDLRPEERFNVRIRHIPESTSRERTLDLRVVLR
                GDIIRVDDLMIRLLEFLKQINNVSLVSIEARTLARAEGDTSIVLVISLRLKIEGEWDE
                SAFQEAVRRVVADLAH"
CDS             complement(join(29415. .29718,29813. .30081,30183. .30443,
                30529. .30656,32107. .32716))
                /note="gene_id:MHM17.8"
                /codon_start=1
                /evidence=not_experimental
                /product="cytokinin oxidase"
                /protein_id="BAA97027.1"
                /db_xref="GI:8777437"
                /translation="MASYNLRSQVRLIAITIVIIITLSTPITTNTSPQPWNILSHNEF
                AGKLTSSSSSVESAATDFGHVTKIFPSAVLIPSSVEDITDLIKLSFDSQLSFPLAARG
                HGHSHRGQASAKDGVVVNMRSMVNRDRGIKVSRTCLYVDVDAAWLWIEVLNKTLELGL
                TPVSWTDYLYLTVGGTLSNGGISGQTFRYGPQITNVLEMDVITGKGEIATCSKDMNSD
                LFFAVLGGLGQFGIITRARIKLEVAPKRAKWLRFLYIDFSEFTRDQERVISKTDGVDF
                LEGSIMVDHGPPDNWRSTYYPPSDHLRIASMVKRHRVIYCLEVVKYYDETSQYTVNEE
                MEELSDSLNHVRGFMYEKDVTYMDFLNRVRTGELNLKSKGQWDVPHPWLNLFVPKTQI
                SKFDDGVFKGIILRNNITSGPVLVYPMNRNKWNDRMSAAIPEEDVFYAVGFLRSAGFD
                NWEAFDQENMEILKFCEDANMGVIQYLPYHSSQEGWVRHFGPRWNIFVERKYKYDPKM
                ILSPGQNIFQKINSS"
tRNA            complement(39239. .39315)
                /product="tRNA-Val"
                /note="codon recognized: GUG; gene_id:MHM17.9"
                /evidence=not_experimental
CDS             complement(join(41045. .41748,41932. .42184))
                /note="unnamed protein product; gene_id:MHM17.10
                pir||T04268
                similar to unknown protein"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="BAA97028.2"

Alignment Scores:
Pred. No.:              1.56e+04        Length:         78423
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x AB024035 (1-78423)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      52493 CATAACCATAACCATAAT 52510

RESULT 499
AC142505
LOCUS       AC142505               78626 bp    DNA     linear   PLN 02-AUG-2004
DEFINITION  Solanum demissum chromosome 5 BAC PGEC287I17 genomic sequence,
            complete sequence.
ACCESSION   AC142505
VERSION     AC142505.1  GI:29469517
KEYWORDS    HTG.
SOURCE      Solanum demissum
  ORGANISM  Solanum demissum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum.
REFERENCE   1  (bases 1 to 78626)
```

```
  AUTHORS   Buell,R., Liu,J., Zaborsky,J., Tallon,L., Wirtz,U., Wei,F.,
            Kuang,H., Zhang,P., Wang,X., Marano,M.R. and Baker,B.
  TITLE     Solanum demissum chromosome 5 BAC PGEC287I17 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 78626)
  AUTHORS   Ronning,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-APR-2003) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 78626)
  AUTHORS   Buell,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUN-2004) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org
REFERENCE   4  (bases 1 to 78626)
  AUTHORS   Buell,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUN-2004) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org
COMMENT     Address all correspondence to:potato@tigr.org

            BAC clone PGEC287I17 is from Solanum demissum chromosome 5
            The orientation of the sequence is from SP6 to T7 end of the BAC
            clone.
            Genes were identified by a combination of several methods: Gene
            prediction programs including Fgenesh (http://www.softberry.com/),
            genscan and Genscan+ (Chris Burge,
            http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHMM (Mark Borodovsky,
            http://genemark.biology.gatech.edu/GeneMark/), and GeneSplicer
            (Mihaela Pertea and Steven Salzberg, contact mpertea@tigr.org),
            searches of the complete sequence against a peptide database and
            the plant EST database at TIGR (http://www.tigr.org/tdb/tgi.shtml).
            Annotated genes are named to indicate the level of evidence for
            their annotation. Genes with similarity to other proteins are named
            after the database hits. Genes without significant peptide
            similarity but with EST similarity are named as  expressed
            proteins. Genes without protein or EST similarity, that are
            predicted by more than two gene prediction programs over most of
            their length are annotated as  hypothetical  proteins. Genes
            encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy,
            http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
            identified by repeatmasker (Arian Smit,
            http://ftp.genome.washington.edu/RM/RepeatMasker.html).
FEATURES             Location/Qualifiers
     source          1. .78626
                     /organism="Solanum demissum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:50514"
                     /chromosome="5"
                     /clone="PGEC287I17"
     repeat_region   187. .288
                     /rpt_family="AT_rich"
     repeat_region   396. .458
                     /rpt_family="AT_rich"
     repeat_region   487. .549
                     /rpt_family="AT_rich"
     gene            complement(1645. .2123)
                     /gene="PGEC287I17.1"
                     /note="predicted by genemarkHMM"
     mRNA            complement(join(<1645. .1710,1793. .1863,1936. .1992,
                     2069. .>2123))
                     /gene="PGEC287I17.1"
     CDS             complement(join(<1645. .1710,1793. .1863,1936. .1992,
                     2069. .2123))
                     /gene="PGEC287I17.1"
                     /codon_start=1
                     /product="hypothetical protein, 3'-partial"
                     /protein_id="AAT39929.1"
                     /db_xref="GI:48057563"
                     /translation="MAGLGIALVPPLQWDEGHSKERPPGDFYSFFTILVAKEGFDLES
                     MVKTTLNWIIGIMGLSRNKRKESGKERFKCLRGFKLGFE"
     repeat_region   2811. .2832
                     /rpt_family="AT_rich"
     repeat_region   3780. .3853
                     /rpt_family="A-rich"
     repeat_region   3857. .3888
                     /rpt_family="AT_rich"
     repeat_region   4291. .4311
                     /rpt_family="AT_rich"
     repeat_region   4340. .4361
                     /rpt_family="AT_rich"
     repeat_region   4610. .4632
                     /rpt_family="AT_rich"
     repeat_region   5234. .5301
                     /rpt_family="AT_rich"
     repeat_region   6069. .6097
                     /rpt_family="AT_rich"
     repeat_region   6196. .6254
                     /rpt_family="AT_rich"
     repeat_region   7438. .7465
                     /rpt_family="AT_rich"
     repeat_region   7995. .8032
                     /rpt_family="AT_rich"
     repeat_region   8732. .8774
                     /rpt_family="AT_rich"
     repeat_region   11063. .11097
                     /rpt_family="AT_rich"
     repeat_region   11824. .11876
                     /rpt_family="(TA)n"
     repeat_region   12934. .12964
                     /rpt_family="AT_rich"
     repeat_region   14818. .14840
                     /rpt_family="AT_rich"
     repeat_region   15264. .15309
                     /rpt_family="AT_rich"
     repeat_region   15674. .15727
                     /rpt_family="AT_rich"
     repeat_region   15814. .15844
                     /rpt_family="AT_rich"
     repeat_region   17879. .18038
                     /rpt_family="(TA)n"
     repeat_region   20198. .20218
                     /rpt_family="AT_rich"
     repeat_region   23693. .23796
                     /rpt_family="(ATGTG)n"
     repeat_region   24946. .24975
                     /rpt_family="AT_rich"
     repeat_region   25068. .25095
                     /rpt_family="AT_rich"
     repeat_region   27972. .28024
                     /rpt_family="AT_rich"
     repeat_region   30074. .30111
                     /rpt_family="AT_rich"
     repeat_region   30477. .30507
                     /rpt_family="AT_rich"
     repeat_region   30539. .30644
                     /rpt_family="AT_rich"
     gene            30890. .35049
                     /gene="PGEC287I17.2"
                     /note="contains Pfam profile  PF00076 (RNA recognition
                     motif); EST BQ120272 from this gene"
     mRNA            join(30890. .30932,32883. .33014,33900. .33975,34414. .34527,
                     34621. .>35049)
                     /gene="PGEC287I17.2"
     CDS             join(32899. .33014,33900. .33975,34414. .34527,34621. .35049)
                     /gene="PGEC287I17.2"
                     /codon_start=1
                     /product="putative RNA recognition motif containing
                     protein"
                     /protein_id="AAT39930.1"
                     /db_xref="GI:48057564"
                     /translation="MNPGGYTAEVTGLSPAATEKDVQEFFGFCGAIEHVEIVRAGEHA
                     STAYVTFRNPHALETAVLLSGAAILDQRVCITSWGHYQDEFDYWNHSSWKPQEDCHSS
                     DSQGHYFVSSAGEAVTLTQDVVKTMLSQGYVLGKGALGKAKAFDESHGLSATAVSKVA
                     DISERIGLTDKFCAGVEVARSVDQRYHISDTTRSAVSATGRTAISAATAVVNSSYFSK
                     GALWMSGALSKAAQAAADLGSRGINK"
```

```
     gene            35853. .45841
                     /gene="PGEC287I17.3"
                     /note="pseudogene, putative sucrose transporter-like
                     protein"
                     /pseudo
     repeat_region   37292. .37426
                     /rpt_family="AT_rich"
     repeat_region   37577. .37623
                     /rpt_family="(TTTTA)n"
     repeat_region   46199. .46250
                     /rpt_family="(TA)n"
     repeat_region   46566. .46595
                     /rpt_family="AT_rich"
     repeat_region   46683. .46708
                     /rpt_family="AT_rich"
     repeat_region   46859. .46879
                     /rpt_family="AT_rich"
     repeat_region   47048. .47095
                     /rpt_family="AT_rich"
     repeat_region   47321. .47354
                     /rpt_family="AT_rich"
     repeat_region   47669. .47735
                     /rpt_family="AT_rich"
     repeat_region   48061. .48155
                     /rpt_family="AT_rich"
     repeat_region   48183. .48211
                     /rpt_family="AT_rich"
     repeat_region   48465. .48552
                     /rpt_family="(TTA)n"
     repeat_region   48826. .48847
                     /rpt_family="(CAATA)n"
     repeat_region   49279. .49308
                     /rpt_family="AT_rich"
     repeat_region   49917. .49940
                     /rpt_family="AT_rich"
     repeat_region   50705. .50732
                     /rpt_family="AT_rich"
     repeat_region   50815. .50835
                     /rpt_family="AT_rich"
     gene            complement(53049. .55667)
                     /gene="PGEC287I17.4"
                     /note="similar to HD zip protein HAHB 1 GB:AAA63765

Alignment Scores:
Pred. No.:              1.57e+04        Length:         78626
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x AC142505 (1-78626)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1450 CACAATCACAATCACAAT 1467

RESULT 500
CR788300
LOCUS       CR788300               78908 bp    DNA     linear   PRI 25-OCT-2004
DEFINITION  Human DNA sequence from clone DAAP-188P13 on chromosome 6, complete
            sequence.
ACCESSION   CR788300
VERSION     CR788300.6  GI:54633117
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 78908)
  AUTHORS   Sycamore,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 25, 2004 this sequence version replaced gi:54606598.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs constructed by
            the MHC Haplotype Consortium and collaborators. Further information
            can be found at
            http://www.sanger.ac.uk/HGP/Chr6/MHC
            DAAP-188P13 is from a APD human bac - B Lymphoblastoid Cell Line
            library VECTOR: pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .78908
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="DAAP-188P13"
                     /clone_lib="APD"
ORIGIN

Alignment Scores:
Pred. No.:              1.57e+04        Length:         78908
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     8               Gaps:           0

US-09-858-332G-28 (1-6) x CR788300 (1-78908)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      62671 CATAATCATAATCATAAT 62688

RESULT 501
AC014989/c
LOCUS       AC014989               79398 bp    DNA     linear   HTG 16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC014989
VERSION     AC014989.1  GI:6436346
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 79398)
  AUTHORS   Adams,M. and Venter,J.C.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10213578 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .79398
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
ORIGIN

Alignment Scores:
Pred. No.:                1.58e+04        Length:         79398
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       14              Gaps:           0

US-09-858-332G-28 (1-6) x AC014989 (1-79398)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      64559 CATAATCATAATCATAAT 64542

RESULT 502
AC139431/c
LOCUS       AC139431               79776 bp    DNA     linear   HTG 03-FEB-2003
DEFINITION  Homo sapiens chromosome 17 clone RP13-680F15 map 17, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC139431
VERSION     AC139431.1  GI:28195962
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 79776)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP13-680F15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 79776)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-FEB-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            ------------- Project Information
                Center project name: L29419
                Center clone name: 680_F_15
            -------------.
            * NOTE: This record contains 89 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      806: contig of 806 bp in length
            *      807      906: gap of 100 bp
            *      907     1707: contig of 801 bp in length
            *     1708     1807: gap of 100 bp
            *     1808     2585: contig of 778 bp in length
            *     2586     2685: gap of 100 bp
            *     2686     3474: contig of 789 bp in length
            *     3475     3574: gap of 100 bp
            *     3575     4376: contig of 802 bp in length
            *     4377     4476: gap of 100 bp
            *     4477     5288: contig of 812 bp in length
            *     5289     5388: gap of 100 bp
            *     5389     6176: contig of 788 bp in length
            *     6177     6276: gap of 100 bp
            *     6277     7069: contig of 793 bp in length
            *     7070     7169: gap of 100 bp
            *     7170     7962: contig of 793 bp in length
            *     7963     8062: gap of 100 bp
            *     8063     8850: contig of 788 bp in length
            *     8851     8950: gap of 100 bp
            *     8951     9748: contig of 798 bp in length
            *     9749     9848: gap of 100 bp
            *     9849    10635: contig of 787 bp in length
            *    10636    10735: gap of 100 bp
            *    10736    11533: contig of 798 bp in length
            *    11534    11633: gap of 100 bp
            *    11634    12431: contig of 798 bp in length
            *    12432    12531: gap of 100 bp
            *    12532    13307: contig of 776 bp in length
            *    13308    13407: gap of 100 bp
            *    13408    14170: contig of 763 bp in length
            *    14171    14270: gap of 100 bp
            *    14271    15067: contig of 797 bp in length
            *    15068    15167: gap of 100 bp
            *    15168    15917: contig of 750 bp in length
            *    15918    16017: gap of 100 bp
            *    16018    16837: contig of 820 bp in length
            *    16838    16937: gap of 100 bp
            *    16938    17739: contig of 802 bp in length
            *    17740    17839: gap of 100 bp
            *    17840    18651: contig of 812 bp in length
            *    18652    18751: gap of 100 bp
            *    18752    19541: contig of 790 bp in length
            *    19542    19641: gap of 100 bp
            *    19642    20440: contig of 799 bp in length
            *    20441    20540: gap of 100 bp
            *    20541    21351: contig of 811 bp in length
            *    21352    21451: gap of 100 bp
            *    21452    22245: contig of 794 bp in length
            *    22246    22345: gap of 100 bp
            *    22346    23146: contig of 801 bp in length
            *    23147    23246: gap of 100 bp
```

```
*    23247    24029: contig of 783 bp in length
*    24030    24129: gap of 100 bp
*    24130    24881: contig of 752 bp in length
*    24882    24981: gap of 100 bp
*    24982    25777: contig of 796 bp in length
*    25778    25877: gap of 100 bp
*    25878    26670: contig of 793 bp in length
*    26671    26770: gap of 100 bp
*    26771    27591: contig of 821 bp in length
*    27592    27691: gap of 100 bp
*    27692    28497: contig of 806 bp in length
*    28498    28597: gap of 100 bp
*    28598    29408: contig of 811 bp in length
*    29409    29508: gap of 100 bp
*    29509    30291: contig of 783 bp in length
*    30292    30391: gap of 100 bp
*    30392    31198: contig of 807 bp in length
*    31199    31298: gap of 100 bp
*    31299    32094: contig of 796 bp in length
*    32095    32194: gap of 100 bp
*    32195    32995: contig of 801 bp in length
*    32996    33095: gap of 100 bp
*    33096    33898: contig of 803 bp in length
*    33899    33998: gap of 100 bp
*    33999    34797: contig of 799 bp in length
*    34798    34897: gap of 100 bp
*    34898    35672: contig of 775 bp in length
*    35673    35772: gap of 100 bp
*    35773    36581: contig of 809 bp in length
*    36582    36681: gap of 100 bp
*    36682    37495: contig of 814 bp in length
*    37496    37595: gap of 100 bp
*    37596    38431: contig of 836 bp in length
*    38432    38531: gap of 100 bp
*    38532    39344: contig of 813 bp in length
*    39345    39444: gap of 100 bp
*    39445    40262: contig of 818 bp in length
*    40263    40362: gap of 100 bp
*    40363    41168: contig of 806 bp in length
*    41169    41268: gap of 100 bp
*    41269    42068: contig of 800 bp in length
*    42069    42168: gap of 100 bp
*    42169    42960: contig of 792 bp in length
*    42961    43060: gap of 100 bp
*    43061    43853: contig of 793 bp in length
*    43854    43953: gap of 100 bp
*    43954    44719: contig of 766 bp in length
*    44720    44819: gap of 100 bp
*    44820    45625: contig of 806 bp in length
*    45626    45725: gap of 100 bp
*    45726    46549: contig of 824 bp in length
*    46550    46649: gap of 100 bp
*    46650    47449: contig of 800 bp in length
*    47450    47549: gap of 100 bp
*    47550    48372: contig of 823 bp in length
*    48373    48472: gap of 100 bp
*    48473    49285: contig of 813 bp in length
*    49286    49385: gap of 100 bp
*    49386    50171: contig of 786 bp in length
*    50172    50271: gap of 100 bp
*    50272    51062: contig of 791 bp in length
*    51063    51162: gap of 100 bp
*    51163    51969: contig of 807 bp in length
*    51970    52069: gap of 100 bp
*    52070    52867: contig of 798 bp in length
*    52868    52967: gap of 100 bp
*    52968    53750: contig of 783 bp in length
*    53751    53850: gap of 100 bp
*    53851    54641: contig of 791 bp in length
*    54642    54741: gap of 100 bp
*    54742    55550: contig of 809 bp in length
*    55551    55650: gap of 100 bp
*    55651    56472: contig of 822 bp in length
*    56473    56572: gap of 100 bp
*    56573    57375: contig of 803 bp in length
*    57376    57475: gap of 100 bp
*    57476    58297: contig of 822 bp in length
*    58298    58397: gap of 100 bp
*    58398    59210: contig of 813 bp in length
*    59211    59310: gap of 100 bp
*    59311    60115: contig of 805 bp in length
*    60116    60215: gap of 100 bp
*    60216    61021: contig of 806 bp in length
*    61022    61121: gap of 100 bp
*    61122    61907: contig of 786 bp in length
*    61908    62007: gap of 100 bp

Alignment Scores:
Pred. No.:                1.59e+04        Length:          79776
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       14              Gaps:            0

US-09-858-332G-28 (1-6) x AC139431 (1-79776)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      63185 CATAATCATAATCATAAT 63168

RESULT 503
AC021665/c
LOCUS       AC021665               80413 bp    DNA     linear   PLN 19-JAN-2001
DEFINITION  Arabidopsis thaliana chromosome 1 BAC F17M19 genomic sequence,
            complete sequence.
ACCESSION   AC021665
VERSION     AC021665.7  GI:12324531
KEYWORDS    HTG.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 80413)
  AUTHORS   Lin,X., Kaul,S., Town,C.D., Benito,M.-I., Creasy,T.H., Haas,B.J.,
            Wu,D., Maiti,R., Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R.,
            Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.
  TITLE     Arabidopsis thaliana chromosome 1 BAC F17M19 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 80413)
  AUTHORS   Lin,X. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JAN-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
REFERENCE   3  (bases 1 to 80413)
  AUTHORS   Town,C.D. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
COMMENT     On Jan 19, 2001 this sequence version replaced gi:12280777.
            Address all correspondence to:at@tigr.org

            BAC clone F17M19 is from Arabidopsis thaliana chromosome 1
            The orientation of the sequence is from SP6 to T7 end of the BAC
            clone.
            Genes were identified by a combination of several methods: Gene
            prediction programs including Genscan+ (Chris Burge,
            http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHMM (Mark Borodovsky,
            http://genemark.biology.gatech.edu/GeneMark/), GlimmerA (a variant
            of GlimmerM, see Mihaela Pertea,
            http://www.tigr.org/softlab/glimmerm_htm/glimmerm.html, and
            GeneSplicer (Mihaela Pertea and Steven Salzberg, contact
            mpertea@tigr.org), searches of the complete sequence against a
            peptide database and the plant EST database at TIGR
            (http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to
```

```
          indicate the level of evidence for their annotation. Genes with
          similarity to other proteins are named after the database hits.
          Genes without significant peptide similarity but with EST
          similarity are named as  unknown  proteins. Genes without protein
          or EST similarity, that are predicted by more than two gene
          prediction programs over most of their length are annotated as
          hypothetical  proteins. Genes encoding tRNAs are predicted by
          tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
          Simple repeats are identified by repeatmasker (Arian Smit,
          http://ftp.genome.washington.edu/RM/RepeatMasker.html).
FEATURES             Location/Qualifiers
     source          1. .80413
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /chromosome="1"
                     /clone="F17M19"
                     /ecotype="Columbia"
     misc_feature    1. .11548
                     /note="overlap with BAC clone F14O23
                     [AC012654:86924. .98471]."
     misc_feature    1
                     /note="32880 nt beyond this point were not included in the
                     submitted sequence due to an overlap with another bac
                     [F14O23]"
     gene            235. .1297
                     /gene="F17M19.1"
                     /note="identical to GB:CAA06978 from [Arabidopsis
                     thaliana]"
     mRNA            join(<235. .369,470. .618,692. .833,930. .1015,1101. .1297)
                     /gene="F17M19.1"
     CDS             join(<235. .369,470. .618,692. .833,930. .1013)
                     /gene="F17M19.1"
                     /codon_start=1
                     /product="protein tyrosine phosphatase, 5'-partial;
                     235-1013"
                     /protein_id="AAG52227.1"
                     /db_xref="GI:12324541"
                     /translation="TVKCGDYFQDEDGPREFGNISLTTKWIKTTDTSLMLRNLEVNYK
                     ETEDQPMSVLHIQYPEWPDHGVPKDTVAVREILKRLYQVPPSLGPIIVHCSAGIGRTG
                     TYCAIHNTIQRILAGDMSALDLAKTVALFRKQRIGMVQTMDQYFFCYNAIVDELEDLT
                     AGTNAGTSS"
     gene            complement(4794. .7233)
                     /gene="F17M19.2"
                     /note="predicted by genefinder"
     mRNA            complement(join(<4794. .5342,6250. .>7233))
                     /gene="F17M19.2"
     CDS             complement(join(4794. .5342,6250. .7233))
                     /gene="F17M19.2"
                     /codon_start=1
                     /product="hypothetical protein; 7233-4794"
                     /protein_id="AAG52224.1"
                     /db_xref="GI:12324538"
                     /translation="MEDKIQSDDFTSHKNPTLPQVIEELKELWAMVLPITAMNCLVYV
                     RAVVSVLFLGRLGSLELAGGALSIGFTNITGYSVMVGLASGLEPVCSQAYGSKNWDLL
                     TLSLHRMVVILLMASLPISLLWINLGPIMLFMGQNPEITATAAEYCLYALPDLLTNTL
                     LQPLRVYLRSQRVTKPMMWCTLAAVAFHVPLNYWLVMVKHWGVPGVAIASVVTNLIMV
                     VLLVGYVWVSGMLQKRVSGDGDGGSTTMVAVVAQSSSVMELVGGLGPLMRVAVPSCLG
                     ICLEWWWYEIVIVMGGYLENPKLAVAATGILIQTTSLMYTVPMALAGCVSARVGNELG
                     AGRPYKARLAANVALACAFVVGALNVAWTVILKERWAGLFTGYEPLKVLVASVMPIVG
                     LCELGNCPQTTGCGILRGTGRPAVGAHVNLGSFYFVGTPVAVGLAFWLKIGFSGLWFG
                     LLSAQAACVVSILYAVLARTDWEGEAVKAMRLTSLEMRKVGQDEESSLLLLDDEKLGD
                     VL"
     gene            26672. .28438
                     /gene="F17M19.3"
                     /note="identical to GB:S38197 from [Arabidopsis thaliana]"
     mRNA            join(<26672. .27931,28018. .28081,28221. .>28438)
                     /gene="F17M19.3"
     CDS             join(26672. .27931,28018. .28081,28221. .28438)
                     /gene="F17M19.3"
                     /codon_start=1
                     /product="sucrose transport protein SUC1; 26672-28438"
                     /protein_id="AAG52225.1"
                     /db_xref="GI:12324539"
                     /translation="MGAYETEKPTKDAAALETQSPEDFDQPSPLRKIISVASIAAGVQ
                     FGWALQLSLLTPYVQLLGIPHKWSSLIWLCGPVSGMIVQPIVGFHSDRCRSKFGRRRP
                     FIATGAALVAVAVFLIGYAADFGYKMGDKLEEKVKVRAIGIFALGFWILDVANNTLQG
                     PCRAFLADLAAGDAKRTRVANAFFSFFMAVGNVLGYAAGSYTNLHKMFPFTMTKACDI
                     YCANLKTCFFLSITLLLIVTVTSLWYVNDKQWSPPPRNADDDEKTSSVPLFGEIFGAF
                     KVMKRPMWMLLIVTALNWIAWFPFLLFDTDWMGREVFGGDSDGNERSKKLYSLGVQSG
                     AMGLMFNSIVLGFMSLGVEWIGRKLGGAKRLWGIVNFILAAGLAMTVLVTKFAEDHRK
                     TAGDLAGPSASVKAGALSLFAVLGIPLAITFSTPFALASIFSSCSGAGQGLSLGVLNL
                     AIVIPQMIVSLGGGPFDALFGGGNLPAFIVAAIAAAISGVLALTVLPSPPPDAPKATT
                     MGGFH"
     gene            30830. .32911
                     /gene="F17M19.4"
                     /note="similar to  sucrose transport protein SUC1
                     GB:S38197 from [Arabidopsis thaliana]"
     mRNA            join(<30830. .32086,32175. .32238,32694. .>32911)
                     /gene="F17M19.4"
     CDS             join(30830. .32086,32175. .32238,32694. .32911)
                     /gene="F17M19.4"
                     /codon_start=1
                     /product="putative sucrose transport protein; 30830-32911"
                     /protein_id="AAG52226.1"
                     /db_xref="GI:12324540"
                     /translation="MGALEAERAANNATALETQSSPEDLGQPSPLRKIISVASIAAGV
                     QFGWALQLSLLTPYIQLLGIPHKWSSYMWLCGPISGMIVQPIVGYHSDRCESRFGRRR
                     PFIAAGVALVAVSVFLIGFAADMGHSFGDKLENKVRTRAIIIFLTGFWFLDVANNTLQ
                     GPCRAFLADLAAGDAKKTRVANACFSFFMAVGNVLGYAAGSYTNLHKMFPFTMTKACD
                     IYCANLKTCFFLSITLLLIVTFSSLWYVKDKQWSPPQGDKEEKTSSLFFFGEIFGAVR
                     HMKRPMVMLLIVTVINWIAWFPFILYDTDWMGREVYGGNSDGDERSKKLYDQGVQAGA
                     LGLMFNSILLGFVSLGVESIGRKMGGAKRLWGCVNFILAIGLAMTVLVTKSAEHHREI
                     AGPLAGPSSGIKAGVFSLFTVLGIPLAITYSIPFALASIFSTNSGAGQGLSLGVLNIA
                     ICIPQMIVSFSSGPLDAQFGGGNLPSFVVGAIAAAVSGVLALTVLPSPPPDAPAMSGA
                     MGFH"
     gene            34092. .36071
                     /gene="F17M19.5"
     mRNA            join(<34092. .34254,34588. .34644,34739. .34829,34933. .35093,
                     35192. .35319,35419. .35556,35672. .35764,35859. .>36071)
                     /gene="F17M19.5"
     CDS             join(34092. .34254,34588. .34644,34739. .34829,34933. .35093,
                     35192. .35319,35419. .35556,35672. .35764,35859. .36071)
                     /gene="F17M19.5"
                     /codon_start=1
                     /product="unknown protein; 34092-36071"
                     /protein_id="AAG52228.1"
                     /db_xref="GI:12324542"
                     /translation="MAESSGSWRDSYKGMSSDNIKGLVLALSSSLFIGASFIVKKKGL
                     KKAASTGTRAGVGGYSYLYEPLWWIGMTTMLLGEIANFAAYAFAPAILVTPLGAVSII
                     ISAVLAHIILREKLHIFGILGCALCVVGSTTIVLHAPQERE IDSVIEVWNLATEPAFM
                     FYASLVIGAAVFLIIRFVPQYGQTNVMVYIGICSLVGSLSVMSVKALGIALKLTFSGT
                     NQLFYPQTWIFTLVVLTCVVTQLNYLNKALDTFNTAIVSPIYYVMFTSLTILASVIMF
                     KDWDRQNGTQIVTEICGFVTILSGTFLLHRTKDMVEGAFLNFIYGTCYAYICSDGLDL
                     EFLHSMRKKRMRI"
     gene            37561. .38082
                     /gene="F17M19.6"
                     /note="predicted by genscan"
     mRNA            <37561. .>38082
                     /gene="F17M19.6"
     CDS             37561. .38082
                     /gene="F17M19.6"
                     /codon_start=1
                     /product="hypothetical protein; 37561-38082"
                     /protein_id="AAG52230.1"
                     /db_xref="GI:12324544"
                     /translation="MALNSVSSEEFTFPSLASHDSSHHSGLDSPPLWKHSPVTSFRRK
                     DYDFGLARLVIGKDGDHDQRKSFSYVEMRSRWIDDKEEEKMDMLWEDLNEELPPRSQS
                     LRIVHGGDGGEKKTALFPDESSAVAVGCGMKLSKKTKKKTNLNVLVLMRVLKKLIVLR
                     SSSQRSPVKTHPR"
     gene            39506. .41337
                     /gene="F17M19.7"
                     /note="similar to histidinol-phosphate aminotransferase
                     GB:CAA70403 from [Nicotiana tabacum]"
     mRNA            join(<39506. .39739,39844. .39931,40054. .40198,40281. .40378,
                     40484. .40592,40683. .40840,40922. .>41337)
```

```
                /gene="F17M19.7"
     CDS        join(39506. .39739,39844. .39931,40054. .40198,40281. .40378,
                40484. .40592,40683. .40840,40922. .41337)
                /gene="F17M19.7"
                /codon_start=1
                /product="putative histidinol-phosphate aminotransferase;
                39506-41337"
                /protein_id="AAG52232.1"
                /db_xref="GI:12324546"
                /translation="MGVINVQGSPSFSIHSSESNLRKSRALKKPFCSIRNRVYCAQSS
                SAAVDESKNITMGDSFIRPHLRQLAAYQPILPFEVLSAQLGRKPEDIVKLDANENPYG

Alignment Scores:
Pred. No.:              1.6e+04         Length:         80413
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x AC021665 (1-80413)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      74673 CACAATCACAATCATAAT 74656

RESULT 504
PFMAL13PA
LOCUS       PFMAL13PA               80518 bp    DNA     linear   HTG 19-AUG-1999
DEFINITION  Plasmodium falciparum 3D7 chromosome 13, *** SEQUENCING IN PROGRESS
            ***, 12 unordered pieces.
ACCESSION   AL109815
VERSION     AL109815.2  GI:5763806
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Plasmodium falciparum 3D7
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1
  AUTHORS   Bowman,S., Churcher,C., Harris,B., Harris,D., Lawson,D., Quail,M.
            and Barrell,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-1999) P.falciparum Genome Sequencing Consortium,
            The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA, UK
COMMENT     On Aug 24, 1999 this sequence version replaced gi:5731892.
            For more information about this sequence or the Malaria Project,
            see http://www.sanger.ac.uk/Projects/P_falciparum. IMPORTANT: This
            sequence is unfinished and does not necessarily represent the
            correct sequence. Work on the sequence is in progress and the
            release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc.
            Order of segments is not known; 800 n's separate segments.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3243: contig of 3243 bp in length
            *     3244     4043: gap of 800 bp
            *     4044     8861: contig of 4818 bp in length
            *     8862     9661: gap of 800 bp
            *     9662    21328: contig of 11667 bp in length
            *    21329    22128: gap of 800 bp
            *    22129    23166: contig of 1038 bp in length
            *    23167    23966: gap of 800 bp
            *    23967    30999: contig of 7033 bp in length
            *    31000    31799: gap of 800 bp
            *    31800    46076: contig of 14277 bp in length
            *    46077    46876: gap of 800 bp
            *    46877    58952: contig of 12076 bp in length
            *    58953    59752: gap of 800 bp
            *    59753    65351: contig of 5599 bp in length
            *    65352    66151: gap of 800 bp
            *    66152    70069: contig of 3918 bp in length
            *    70070    70869: gap of 800 bp
            *    70870    74298: contig of 3429 bp in length
            *    74299    75098: gap of 800 bp
            *    75099    77479: contig of 2381 bp in length
            *    77480    78279: gap of 800 bp
            *    78280    80518: contig of 2239 bp in length.
FEATURES            Location/Qualifiers
     source         1. .80518
                    /organism="Plasmodium falciparum 3D7"
                    /mol_type="genomic DNA"
                    /strain="3D7"
                    /isolate="3D7"
                    /db_xref="taxon:36329"
                    /chromosome="13"
ORIGIN

Alignment Scores:
Pred. No.:              1.6e+04         Length:         80518
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x PFMAL13PA (1-80518)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      39802 CACAATCATAATCATAAT 39819

RESULT 505
AP004577
LOCUS       AP004577                80655 bp    DNA     linear   PLN 22-JUL-2003
DEFINITION  Lotus corniculatus var. japonicus genomic DNA, chromosome 2,
            clone:LjT10N22, TM0381, complete sequence.
ACCESSION   AP004577
VERSION     AP004577.1  GI:21624382
KEYWORDS    HTG.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1
  AUTHORS   Stracke,S., Kistner,C., Yoshida,S., Mulder,L., Sato,S., Kaneko,T.,
            Tabata,S., Sandal,N., Stougaard,J., Szczyglowski,K. and Parniske,M.
  TITLE     A plant receptor-like kinase required for both bacterial and fungal
            symbiosis
  JOURNAL   Nature 417 (6892), 959-962 (2002)
   PUBMED   12087405
REFERENCE   2
  AUTHORS   Sato,S., Kaneko,T., Nakamura,Y., Asamizu,E., Kato,T. and Tabata,S.
  TITLE     Structural Analysis of a Lotus japonicus Genome
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 80655)
  AUTHORS   Sato,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-DEC-2001) Shusei Sato, Kazusa DNA Research Institute,
            The First Laboratory for Plant Gene Research; 2-6-7
            kazusa-kamatari, Kisarazu, Chiba 292-0812, Japan
            (E-mail:ssato@kazusa.or.jp, URL:http://www.kazusa.or.jp/,
            Tel:81-438-52-3935(ex.2337), Fax:81-438-52-3934)
FEATURES            Location/Qualifiers
     source         1. .80655
                    /organism="Lotus corniculatus var. japonicus"
                    /mol_type="genomic DNA"
```

```
                     /variety="japonicus"
                     /db_xref="taxon:34305"
                     /chromosome="2"
                     /clone="LjT10N22"
                     /clone_lib="LjT library"
                     /note="TAC clone:TM0381
                     synonym: Lotus japonicus"
ORIGIN

Alignment Scores:
Pred. No.:                1.6e+04        Length:         80655
Score:                    42.00          Matches:        6
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       15             Gaps:           0

US-09-858-332G-28 (1-6) x AP004577 (1-80655)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      64080 CACAACCACAACCACAAC 64097

RESULT 506
AC165439
LOCUS       AC165439               81250 bp    DNA     linear   HTG 04-AUG-2005
DEFINITION  Medicago truncatula clone mth2-146b13, WORKING DRAFT SEQUENCE, 26
            unordered pieces.
ACCESSION   AC165439
VERSION     AC165439.5  GI:71795771
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 81250)
  AUTHORS   Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
            and Roe,B.A.
  TITLE     Medicago truncatula BAC Clone mth2-146b13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 81250)
  AUTHORS   Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
            and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUL-2005) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 81250)
  AUTHORS   Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
            and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-AUG-2005) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Aug 4, 2005 this sequence version replaced gi:71725507.
            -------------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ----------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2038: contig of 2038 bp in length
            *     2039     2138: gap of unknown length
            *     2139     4145: contig of 2007 bp in length
            *     4146     4245: gap of unknown length
            *     4246     6438: contig of 2193 bp in length
            *     6439     6538: gap of unknown length
            *     6539     8810: contig of 2272 bp in length
            *     8811     8910: gap of unknown length
            *     8911    11503: contig of 2593 bp in length
            *    11504    11603: gap of unknown length
            *    11604    13901: contig of 2298 bp in length
            *    13902    14001: gap of unknown length
            *    14002    16632: contig of 2631 bp in length
            *    16633    16732: gap of unknown length
            *    16733    19086: contig of 2354 bp in length
            *    19087    19186: gap of unknown length
            *    19187    21589: contig of 2403 bp in length
            *    21590    21689: gap of unknown length
            *    21690    23736: contig of 2047 bp in length
            *    23737    23836: gap of unknown length
            *    23837    25836: contig of 2000 bp in length
            *    25837    25936: gap of unknown length
            *    25937    28858: contig of 2922 bp in length
            *    28859    28958: gap of unknown length
            *    28959    31556: contig of 2598 bp in length
            *    31557    31656: gap of unknown length
            *    31657    33791: contig of 2135 bp in length
            *    33792    33891: gap of unknown length
            *    33892    36006: contig of 2115 bp in length
            *    36007    36106: gap of unknown length
            *    36107    38990: contig of 2884 bp in length
            *    38991    39090: gap of unknown length
            *    39091    41804: contig of 2714 bp in length
            *    41805    41904: gap of unknown length
            *    41905    44262: contig of 2358 bp in length
            *    44263    44362: gap of unknown length
            *    44363    47958: contig of 3596 bp in length
            *    47959    48058: gap of unknown length
            *    48059    51269: contig of 3211 bp in length
            *    51270    51369: gap of unknown length
            *    51370    54970: contig of 3601 bp in length
            *    54971    55070: gap of unknown length
            *    55071    58944: contig of 3874 bp in length
            *    58945    59044: gap of unknown length
            *    59045    63785: contig of 4741 bp in length
            *    63786    63885: gap of unknown length
            *    63886    68788: contig of 4903 bp in length
            *    68789    68888: gap of unknown length
            *    68889    74365: contig of 5477 bp in length
            *    74366    74465: gap of unknown length
            *    74466    81250: contig of 6785 bp in length.
FEATURES             Location/Qualifiers
     source          1. .81250
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3880"
                     /clone="mth2-146b13"
                     /clone_lib="Medicago truncatula BAC library H2"
     gap             2039. .2138
                     /estimated_length=unknown
     gap             4146. .4245
                     /estimated_length=unknown
     gap             6439. .6538
                     /estimated_length=unknown
     gap             8811. .8910
                     /estimated_length=unknown
     gap             11504. .11603
                     /estimated_length=unknown
     gap             13902. .14001
                     /estimated_length=unknown
     gap             16633. .16732
                     /estimated_length=unknown
     gap             19087. .19186
                     /estimated_length=unknown
     gap             21590. .21689
                     /estimated_length=unknown
```

```
     gap             23737. .23836
                     /estimated_length=unknown
     gap             25837. .25936
                     /estimated_length=unknown
     gap             28859. .28958
                     /estimated_length=unknown
     gap             31557. .31656
                     /estimated_length=unknown
     gap             33792. .33891
                     /estimated_length=unknown
     gap             36007. .36106
                     /estimated_length=unknown
     gap             38991. .39090
                     /estimated_length=unknown
     gap             41805. .41904
                     /estimated_length=unknown
     gap             44263. .44362
                     /estimated_length=unknown
     gap             47959. .48058
                     /estimated_length=unknown
     gap             51270. .51369
                     /estimated_length=unknown
     gap             54971. .55070
                     /estimated_length=unknown
     gap             58945. .59044
                     /estimated_length=unknown
     gap             63786. .63885
                     /estimated_length=unknown
     gap             68789. .68888
                     /estimated_length=unknown
     gap             74366. .74465
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              1.62e+04        Length:         81250
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AC165439 (1-81250)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       9399 CATAATCATAATCATAAT 9416

RESULT 507
CR382329
LOCUS       CR382329               81358 bp    DNA     linear   VRT 12-JAN-2005
DEFINITION  Zebrafish DNA sequence from clone DKEY-72L19 in linkage group 8,
            complete sequence.
ACCESSION   CR382329
VERSION     CR382329.10  GI:57635821
KEYWORDS    HTG.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 81358)
  AUTHORS   Glithero,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jan 13, 2005 this sequence version replaced gi:57506454.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived
            Zebrafish pUC subclones occasionally display inconsistency over the
            length of mononucleotide A/T runs and conserved TA repeats. Where
            this is found the longest good quality representation will be
            submitted.
            Repeat names beginning 'Dr' were identified by the Recon repeat
            discovery system (Zhirong Bao and Sean Eddy, submitted), and those
            beginning 'drr' were identified by Rick Waterman (Stephen Johnson
            lab, WashU).  For further information see
            http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml DKEY-72L19
            is from a Zebrafish BAC library
            VECTOR: pIndigoBAC-5.
FEATURES             Location/Qualifiers
     source          1. .81358
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-72L19"
                     /clone_lib="DanioKey"
ORIGIN

Alignment Scores:
Pred. No.:              1.62e+04        Length:         81358
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     5               Gaps:           0

US-09-858-332G-28 (1-6) x CR382329 (1-81358)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       3174 CATAATCATAACCATAAT 3191

RESULT 508
AP006384/c
LOCUS       AP006384               81705 bp    DNA     linear   PLN 22-JUL-2003
DEFINITION  Lotus corniculatus var. japonicus genomic DNA, chromosome 1,
            clone:LjT10I22, TM0248, complete sequence.
ACCESSION   AP006384
VERSION     AP006384.1  GI:31581015
KEYWORDS    HTG.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1
  AUTHORS   Asamizu,E., Kato,T., Sato,S., Nakamura,Y., Kaneko,T. and Tabata,S.
  TITLE     Structural Analysis of a Lotus japonicus Genome. IV. Sequence
            Features and Mapping of seventy-three TAC clones which cover the
            7.5 Mb Regions of the Genome
```

```
  JOURNAL   DNA Res. (2003) In press
REFERENCE   2  (bases 1 to 81705)
  AUTHORS   Sato,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAY-2003) Shusei Sato, Kazusa DNA Research Institute,
            Department of Plant Gene Research; 2-6-7 Kazusa-kamatari, Kisarazu,
            Chiba 292-0818, Japan (E-mail:ssato@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/, Tel:81-438-52-3935(ex.2337),
            Fax:81-438-52-3934)
FEATURES             Location/Qualifiers
     source          1. .81705
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="genomic DNA"
                     /variety="japonicus"
                     /db_xref="taxon:34305"
                     /chromosome="1"
                     /clone="LjT10I22"
                     /clone_lib="LjT library"
                     /note="TAC clone:TM0248
                     synonym: Lotus japonicus"
ORIGIN

Alignment Scores:
Pred. No.:                1.62e+04        Length:         81705
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       15              Gaps:           0

US-09-858-332G-28 (1-6) x AP006384 (1-81705)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      79931 CATAATCATAATCATAAT 79914

RESULT 509
AP008165/c
LOCUS       AP008165               82418 bp    DNA     linear   HTG 28-DEC-2004
DEFINITION  Lotus corniculatus var. japonicus clone LjT15F02, *** SEQUENCING IN
            PROGRESS ***, 17 unordered pieces.
ACCESSION   AP008165
VERSION     AP008165.1  GI:56806471
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1
  AUTHORS   Kaneko,T., Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Structural Analysis of a Lotus japonicus Genome. XI. Sequence
            Features and Mapping of Nine hundred twenty-one TAC Clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 82418)
  AUTHORS   Sato,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-2004) Shusei Sato, Kazusa DNA Research Institute,
            Department of Plant Gene Research; 2-6-7 Kazusa-kamatari, Kisarazu,
            Chiba, 292-0818, Japan (E-mail:ssato@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/, Tel:81-438-52-3935(ex.2337),
            Fax:81-438-52-3934)
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1496: contig of 1496 bp in length
            *     1497     1596: gap of unknown length
            *     1597     3149: contig of 1553 bp in length
            *     3150     3249: gap of unknown length
            *     3250     3742: contig of 493 bp in length
            *     3743     3842: gap of unknown length
            *     3843     5991: contig of 2149 bp in length
            *     5992     6091: gap of unknown length
            *     6092     7414: contig of 1323 bp in length
            *     7415     7514: gap of unknown length
            *     7515     9811: contig of 2297 bp in length
            *     9812     9911: gap of unknown length
            *     9912    12477: contig of 2566 bp in length
            *    12478    12577: gap of unknown length
            *    12578    15022: contig of 2445 bp in length
            *    15023    15122: gap of unknown length
            *    15123    17526: contig of 2404 bp in length
            *    17527    17626: gap of unknown length
            *    17627    20955: contig of 3329 bp in length
            *    20956    21055: gap of unknown length
            *    21056    24542: contig of 3487 bp in length
            *    24543    24642: gap of unknown length
            *    24643    28253: contig of 3611 bp in length
            *    28254    28353: gap of unknown length
            *    28354    36975: contig of 8622 bp in length
            *    36976    37075: gap of unknown length
            *    37076    44097: contig of 7022 bp in length
            *    44098    44197: gap of unknown length
            *    44198    55301: contig of 11104 bp in length
            *    55302    55401: gap of unknown length
            *    55402    78905: contig of 23504 bp in length
            *    78906    79005: gap of unknown length
            *    79006    82418: contig of 3413 bp in length.
FEATURES             Location/Qualifiers
     source          1. .82418
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="genomic DNA"
                     /variety="japonicus"
                     /db_xref="taxon:34305"
                     /clone="LjT15F02"
                     /clone_lib="LjT library"
                     /note="TAC clone:TM1582, synonym:Lotus japonicus"
     gap             1497. .1596
                     /estimated_length=unknown
     gap             3150. .3249
                     /estimated_length=unknown
     gap             3743. .3842
                     /estimated_length=unknown
     gap             5992. .6091
                     /estimated_length=unknown
     gap             7415. .7514
                     /estimated_length=unknown
     gap             9812. .9911
                     /estimated_length=unknown
     gap             12478. .12577
                     /estimated_length=unknown
     gap             15023. .15122
                     /estimated_length=unknown
     gap             17527. .17626
                     /estimated_length=unknown
     gap             20956. .21055
                     /estimated_length=unknown
     gap             24543. .24642
                     /estimated_length=unknown
     gap             28254. .28353
                     /estimated_length=unknown
     gap             36976. .37075
                     /estimated_length=unknown
     gap             44098. .44197
                     /estimated_length=unknown
     gap             55302. .55401
                     /estimated_length=unknown
     gap             78906. .79005
                     /estimated_length=unknown
ORIGIN
```

```
Alignment Scores:
Pred. No.:              1.64e+04        Length:         82418
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AP008165 (1-82418)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      70060 CACAACCACAACCATAAC 70043

RESULT 510
ATT8O5
LOCUS       ATT8O5                 83029 bp    DNA     linear   PLN 16-APR-2005
DEFINITION  Arabidopsis thaliana DNA chromosome 4, BAC clone T8O5 (ESSAII
            project).
ACCESSION   AL021890
VERSION     AL021890.1  GI:2894557
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Bevan,M., Wedler,H., Wambutt,R., Bancroft,I., Mewes,H.W., Mayer,K.
            and Schueller,C.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 83029)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-FEB-1998) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
FEATURES             Location/Qualifiers
     source          1. .83029
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /chromosome="4"
                     /ecotype="Columbia"
     gene            1. .1440
                     /gene="T8O5.10"
     CDS             join(1. .79,179. .459,515. .805,893. .1098,1176. .1440)
                     /gene="T8O5.10"
                     /note="Protein sequence is in conflict with the conceptual
                     translation; similarity to Hypothetical protein,
                     Saccharomyces cerevisiae, MNOS:S71693"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAA17147.1"
                     /db_xref="GI:2894558"
                     /db_xref="GOA:O49703"
                     /db_xref="InterPro:IPR004130"
                     /db_xref="UniProt/TrEMBL:O49703"
                     /translation="KLRVSAASSSSNFKKKPIIIIVVGMAGSGKTSFLHRLVCHTFDS
                     KSHGYVVNLDPAVMSLPFGANIDIRDTVKYKEVMKQYNLGPNGGILTSLNLFATKFDE
                     VNIPLELVQMITDFFSHCVVSVIEKRADQLDYVLVDTPGQIEIFTWSASGAIITEAFA
                     STFPTVVTYVVDTPRSSSPITFMSNMLYACSILYKTRLPLVLAFNKTDVADHKFALEW
                     MEDFEVFQAAIQSDNSYTATLANSLSLSLYEFYRNIRSVGVSAISGAGMDGFFKAIEA
                     SAEEYMETYKADLDMRKADKERLEEERKKHEMEKLRKDMESSQGGTVVLNTGLKDRDA
                     TEKMMLEEDDEDFQVEDEEDSDDAIDEDDEDDETKHYYL"
     exon            1. .79
                     /gene="T8O5.10"
                     /number=1
     intron          80. .178
                     /gene="T8O5.10"
                     /number=1
     exon            179. .459
                     /gene="T8O5.10"
                     /number=2
     intron          460. .514
                     /gene="T8O5.10"
                     /number=2
     exon            515. .805
                     /gene="T8O5.10"
                     /number=3
     intron          806. .892
                     /gene="T8O5.10"
                     /number=3
     exon            893. .1098
                     /gene="T8O5.10"
                     /number=4
     intron          1099. .1175
                     /gene="T8O5.10"
                     /number=4
     exon            1176. .1440
                     /gene="T8O5.10"
                     /number=5
     gene            1809. .4079
                     /gene="T8O5.20"
     gene            complement(join(1809. .1991,2084. .2137,2228. .2311,
                     2409. .2623,3028. .3127,4038. .4079))
                     /gene="T8O5.20"
     CDS             complement(join(1809. .1991,2084. .2137,2228. .2311,
                     2409. .2623,3028. .3127,4038. .4079))
                     /gene="T8O5.20"
                     /note="similarity to NADH dehydrogenase chain 4,
                     Acanthamoeba castellanii mitochondrion, PIR2:S53834
                     contains EST gb:N38447, T21918"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAA17148.1"
                     /db_xref="GI:2894559"
                     /db_xref="InterPro:IPR007599"
                     /db_xref="UniProt/TrEMBL:O49704"
                     /translation="MTPDYVAAPVKRTAIISPYNLYLNPTLVVKQYQFWRLVTNFLYF
                     RKMDLDFLFHMFFLARYCKLLEENSFRGKTTDFLYMLLFGATVLTGIVLIGGMIPYLS
                     VSFSKIIFLSNSLTFMMVYVWSKQNPYIHMSFLGLFTFTAAYLPWVLLGFSILVGASA
                     WGDFLGMIAGHAYYFLAFVYPRMTDRRPLKTPSFLKALFADEPVVIARPEDVRFAHAP
                     FDEIHQD"
     exon            complement(1809. .1991)
                     /gene="T8O5.20"
                     /number=1
     intron          complement(1992. .2083)
                     /gene="T8O5.20"
                     /number=1
     exon            complement(2084. .2137)
                     /gene="T8O5.20"
                     /number=2
     intron          complement(2138. .2227)
                     /gene="T8O5.20"
                     /number=2
     exon            complement(2228. .2311)
                     /gene="T8O5.20"
                     /number=3
     intron          complement(2312. .2408)
                     /gene="T8O5.20"
                     /number=3
     exon            complement(2409. .2623)
                     /gene="T8O5.20"
                     /number=4
     intron          complement(2624. .3027)
                     /gene="T8O5.20"
                     /number=4
     exon            complement(3028. .3127)
                     /gene="T8O5.20"
                     /number=5
     intron          complement(3128. .4037)
                     /gene="T8O5.20"
```

```
                /number=5
     gene       4022. .10335
                /gene="T8O5.30"
     CDS        join(4022. .4729,4822. .5290,5735. .5961,6060. .6125,
                6311. .6364,6732. .6848,7056. .7154,7211. .7391,7599. .7736,
                7938. .8132,8214. .8494,9384. .9606,9673. .9937,9981. .10136,
                10248. .10335)
                /gene="T8O5.30"
                /note="similarity to microtubule associated protein,
                Drosophila melanogaster, PATCHX:G1930122"
                /codon_start=1
                /product="putative protein"
                /protein_id="CAA17149.1"
                /db_xref="GI:2894560"
                /db_xref="InterPro:IPR000048"
                /db_xref="InterPro:IPR000225"
                /db_xref="UniProt/TrEMBL:O49705"
                /translation="MDENEPPCASPAPPRNPASSLLSDISNFKTPRRTSVVNSNISKS
                PYPHFFTASKQTPKSSSSNFRRPSMVHSYASRSKVSTSSRRLKAFELQQSQSSRKAEL
                TKEKNLRSLAKSLTVWLNFLFENPENCGCDPFENESGVGNLGKGKRDSGEALGNSKSV
                GVDTMWRSPKRLRNLGWCGEKKRSEIDSSLTGSKYSTLRESLRDVCSLDDLKQRMQFH
                LSLGSCKEIFDVMTRVSKNIDEGRIKMKPQCPLVTDFGMKEKAIKALMSYNQVWLRLG
                LYIIFGGDSFLSDSEVNSDQEMAFLKMVISKQFFSHDGLARSYVYNKMVEGLYRPGYY
                EALGNVILKRILLLVLVIDRAKSQSCLSLKYGIDGIDGGSPLMFSEKSSIKSSHQLIC
                GSEFDLMRVLIKQKMVVPSDNRKKNLANCRIALQYLKDAGVSLKDDEGMMITGEDVAD
                GDRELTISLLWNIFVHLQLPLLINGRLLTEEIYKVQGVEQNNQITMSTPLEMLLNWIQ
                DPGGQEGPQSVMSNTDYHDAVQNFILSQKLTALLGSFPEIGDLLEHNAVVSNQSVIIL
                LAFLSSKLIVKENLLDFTLAGYLTNSVFSQEQLNFHKLLCSSCQVQEKRYSRIRISCS
                SSEAVTIEEPDRENGEDATKRFQAIKAWWQDMANQNQISVGKANSHTLQGSLSRKCTT
                DSQRAYLLLAEIAAVIIQSNIRGLHARRKFRNKMRAICFLQAAVRTWLSVKNIQVVEK
                FNVEEVTLHLSERSANLKPVARYVKFIVDRSRFIKLRKSVSVIQKAVRRHQSNLHHEL
                KAALRIQLAWRSYKEKVISSITIQSYVRGWITRRMNRTYKFSSILIQFQVSFYPYISQ
                SHWRGYLTRKASKAQVLDLRVRMQTSAANIDDKKRLINKLLSALSELLSMKNVHNILH
                ICETLESATKYSDKCCEELVAAGAIEKLLTLIRSASRSIPDQQVSKHALSTLGHLARY
                PQMADELINTKGSIQTIFWELLRSSFSSQVFFLSFKNKEEAYFIASDVLKKICSSHKG
                VEAVRKLPALVKRLHVLVEELTRKANIEKRNVKGQTGKEKSERRLKEAIELVKLITRR
                "
     exon       4022. .4729
                /gene="T8O5.30"
                /number=1
     exon       complement(4038. .4079)
                /gene="T8O5.20"
                /number=6
     intron     4730. .4821
                /gene="T8O5.30"
                /number=1
     exon       4822. .5290
                /gene="T8O5.30"
                /number=2
     intron     5291. .5734
                /gene="T8O5.30"
                /number=2
     exon       5735. .5961
                /gene="T8O5.30"
                /number=3
     intron     5962. .6059
                /gene="T8O5.30"
                /number=3
     exon       6060. .6125
                /gene="T8O5.30"
                /number=4
     intron     6126. .6310
                /gene="T8O5.30"
                /number=4
     exon       6311. .6364
                /gene="T8O5.30"
                /number=5
     intron     6365. .6731
                /gene="T8O5.30"
                /number=5
     exon       6732. .6848

Alignment Scores:

Pred. No.:              1.65e+04        Length:         83029
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x ATT8O5 (1-83029)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       3476 CACAATCACAATCACAAT 3493

RESULT 511
ATM4I22/c
LOCUS       ATM4I22                83371 bp    DNA     linear   PLN 16-APR-2005
DEFINITION  Arabidopsis thaliana DNA chromosome 4, P1 clone M4I22 (ESSA
            project).
ACCESSION   AL030978
VERSION     AL030978.1  GI:3269280
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-2000) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
FEATURES             Location/Qualifiers
     source          1. .83371
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /chromosome="4"
                     /ecotype="Columbia"
     LTR             540. .784
     gene            737. .3246
                     /gene="AT4g27200"
     CDS             join(737. .1811,1828. .2002,2037. .3246)
                     /gene="AT4g27200"
                     /note="similarity to retrotransposon Opie-2, Zea mays,
                     PID:g1657768
                     contains EST gb:T46634, AA404771, AI997526.1"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAA19714.1"
                     /db_xref="GI:3269281"
                     /db_xref="UniProt/TrEMBL:O81823"
                     /translation="MLVRKRHSTSSWRIGSACFPTLRDYAENKFNPCSLKCVFLGYNE
                     KYKGYRCLYPPTGRLYISRHVIFDESVYPFSHTYKHLHPQPRTPLLAAWLRSSDSPAP
                     STSTSPSSRSPLFTSADFPPLPQRKTPLLPTLVPISSVSHASNITTQQSPDFDSERTT
                     DFDSASIGDSSHSSQAGSDSEETIQQASVNVHQTPASTNVHPMVTRAKVGISKPNPRY
                     VFLSHKVSYPEPKTVTAALKHPGWTGAMTEEIGNCSETQTWSLVPYKSDMHVLGSKWV
                     FRTKLHADGTLNKLKARIVAKGFLQEEGIDYLETYSPVVRTPTVRLVLHLATALNWDI
                     KQMDVKNAFLHGDLKETVYMTQPAANRDHVCLLHKSIYGLKQSPRAWFDKFSTFLLEF
                     GFFCRKSDPSLFIYAHNNNLILLLLSQTLTSLLAALNKEFRMTDMGQHSLTFLGIQVQ
                     RQQNGLFMSQQKYAEDLLIAASMEHCTPLPTPLPVQLDRVPHQEELFSDPTYFRSIAG
                     KLQYLTLTRPDIQFAVNFVCQKMHQPTISDFHLLKRILRYIKGTITMGISYSRDSPTL
                     LQAYSDSDWGNCKQTRRSVGGLCTFMGTNLVSWSSKKHPTVSRSSTEAEYKSLSDAAS
                     EILWLSTLLRELRIPLPDTPELFCDNLSAVYLTANPAFHARTKHFDIDFHFVRERVAL
                     KALVVKHIPGSEQIADIFTKSLPYEAFIHLRGKLGVTLPPTPSLRGTINTIQVKPRQN
                     FVVSSNSAKKQIKAQCRPKKPTTQVWKQRSSETKTETPEASPSHDKNRAGTTSPNACT
                     EKAFNIKLENRFLTLDSEIPA"
     exon            737. .1811
                     /gene="AT4g27200"
                     /number=1
```

```
intron          1812. .1827
                /gene="AT4g27200"
                /number=1
exon            1828. .2002
                /gene="AT4g27200"
                /number=2
intron          2003. .2036
                /gene="AT4g27200"
                /number=2
exon            2037. .3246
                /gene="AT4g27200"
                /number=3
LTR             2116. .3372
LTR             2973. .3217
                /gene="AT4g27200"
gene            3583. .7931
                /gene="AT4g27210"
CDS             join(3583. .3719,3767. .4131,4357. .4828,4949. .7931)
                /gene="AT4g27210"
                /note="similarity to Hopscotch retrotransposon, Zea mays,
                PID:g531389
                contains EST gb:AW004308.1, AA712235, Z35393"
                /codon_start=1
                /product="putative protein"
                /protein_id="CAA19715.1"
                /db_xref="GI:3269282"
                /db_xref="GOA:O81824"
                /db_xref="InterPro:IPR001584"
                /db_xref="UniProt/TrEMBL:O81824"
                /translation="MGQPLLHQIRSTKPGSRQTESSNLGFLDPFPRISSVWLSTATLH
                MRLFELQRRLQNVSKRDKTMDAYLNDLKNICDQLASVGSPVTEKMKIFAALNGLGREY
                EPIKTTIENSMDTQPGPSLENVIPKLTGYDDRLQGYLEETTISPYVAFNITTSDDSYR
                NSYNRGKDVTDQHGNEWLPDSAATAHVTNSPRSLQQSQPYHGTDAIMVDDGNYLPITH
                TGSTNLASSSGTVPLTDVLVCPSITKSLLSMSKLTQDFPCTVEPEYDGVRVNDKATKK
                LLLMGSNRDGLYCLKDDKQFQAFFSTRQRSASDEVWHRRLGHPHPQILQPLERVHCDL
                WGPTTITSVQGFRYYAVFIDHYSRFSWIYPLKLKSDFYNIFLAFHKLVENQLSQKISV
                FQCDGGGEFVSHKFLQHLQSHGIQQQLSCPHTPQQNGLAERKHRHLVELGLSMLFQSH
                VPHKFWVEAFFTANFLINLLPTSALKESISPYEKLYDKKPDYTSLRSFGSACFPTLRD
                YAENKFNPCSLKCVFLGYNEKYKGYRCLYPPTGRLYISRHVIFDESVYPFSHTYKHLH
                PQPRTPLLAAWLRSSDSPAPSTSTSPSSRSPLFTSADFPPLPQRKTPLLPTLVPISSV
                SHASNITTQQSPDFDSERTTDFDSASIGDSSHSSQAGSDSEETIQQASVNVHQTHAST
                NVHPMVTRAKVGISKPNPRYVFLSHKVSYPEPKTVTAALKHPGWTGAMTEEIGNCSET
                QTWSLVPYKSDMHVLGSKWVFRTKLHADGTLNKLKARIVAKGFLQEEGIDYLETYSPV
                VRTPTVRLVLHLATALNWDIKQMDVKNAFLHGDLKETVYMTQPAGFVDPSKPDHVCLL
                HKSIYGLKQSPRAWFDKFSTFLLEFGFFCSKSDPSLFIYAHNNNLILLLLYVDDMVIT
                GNSSQTLTSLLAALNKEFRMTDMGQLHYFLGIQVQRQQNGLFMSQQKYAEDLLIAASM
                EHCTPLPTPLPVQLDRVPHQEELFSDPTYFRSIAGKLQYLTLTRPDIQFAVNFVCQKM
                HQPTISDFHLLKRILRYIKGTITMGISYSRDSPTLLQAYSDSDWGNCKQTRRSVGGLC
                TFMGTNLVSWSSKKHPTVSRSSTEAEYKSLSDAASEILWLSTLLRELRIPLPDTPELF
                CDNLSAVYLTANPAFHARTKHFDIDFHFVRERVALKALVVKHIPGSEQIADIFTKSLP
                YEAFIHLRGKLGVTLSPTPSLRGTINTIQVKPRQNFVVSSNSAKKQMKAQCRPKKPTT
                QVWKQRSSAIKTETPEVSPSYDKNRAGTTSPNACTEKAFNIKLENRFLTLDSEIPA"
exon            3583. .3719
                /gene="AT4g27210"
                /number=1
intron          3720. .3766
                /gene="AT4g27210"
                /number=1
exon            3767. .4131
                /gene="AT4g27210"
                /number=2
intron          4132. .4356
                /gene="AT4g27210"
                /number=2
exon            4357. .4828
                /gene="AT4g27210"
                /number=3
intron          4829. .4948
                /gene="AT4g27210"
                /number=3
exon            4949. .7931
                /gene="AT4g27210"
                /number=4
LTR             6801. .8057
gene            9962. .12721
                /gene="AT4g27220"
gene            complement(9962. .12721)
                /gene="AT4g27220"
CDS             complement(9962. .12721)
                /gene="AT4g27220"
                /note="similarity to disease resistance protein RPS2,
                Arabidopsis thaliana, PIR2:A54809
                Contains ATP/GTP-binding site motif A (P-loop)
                AA141-148;Prokaryotic membrane lipoprotein lipid
                attachment site AA750-760;Eukaryotic thiol (cysteine)
                proteases active sites AA449-459"
                /codon_start=1
                /product="putative protein"
                /protein_id="CAA19716.1"
                /db_xref="GI:3269283"
                /db_xref="GOA:O81825"
                /db_xref="InterPro:IPR000219"
                /db_xref="InterPro:IPR000767"
                /db_xref="InterPro:IPR001611"
                /db_xref="InterPro:IPR002182"
                /db_xref="InterPro:IPR003591"
                /db_xref="UniProt/Swiss-Prot:O81825"
                /translation="MFRSNARALNRALERLKNVQTKVNEALKRSGIQEKSLERKLRIW
                LRKVEENVPLGELILEKRSSCAIWLSDKDVEILEKVKRLEEQGQDLIKKISVNKSSRE
                IVERVLGPSFHPQKTALEMLDKLKDCLKKKNVQKIGVWGMGGVGKTTLVRTLNNDLLK
                YAATQQFALVIWVTVSKDFDLKRVQMDIAKRLGKRFTREQMNQLGLTICERLIDLKNF
                LLILDDVWHPIDLDQLGIPLALERSKDSKVVLTSRRLEVCQQMMTNENIKVACLQEKE
                AWELFCHNVGEVANSDNVKPIAKDVSHECCGLPLAIITIGRTLRGKPQVEVWKHTLNL
                LKRSAPSIDTEEKIFGTLKLSYDFLQDNMKSCFLFCALFPEDYSIKVSELIMYWVAEG
                LLDGQHHYEDMMNEGVTLVERLKDSCLLEDGDSCDTVKMHDVVRDFAIWFMSSQGEGF
                HSLVMAGRGLIEFPQDKFVSSVQRVSLMANKLERLPNNVIEGVETLVLLLQGNSHVKE
                VPNGFLQAFPNLRILDLSGVRIRTLPDSFSNLHSLRSLVLRNCKKLRNLPSLESLVKL
                QFLDLHESAIRELPRGLEALSSLRYICVSNTYQLQSIPAGTILQLSSLEVLDMAGSAY
                SWGIKGEEREGQATLDEVTCLPHLQFLAIKLLDVLSFSYEFDSLTKRLTKFQFLFSPI
                RSVSPPGTGEGCLAISDVNVSNASIGWLLQHVTSLDLNYCEGLNGMFENLVTKSKSSF
                VAMKALSIHYFPSLSLASGCESQLDLFPNLEELSLDNVNLESIGELNGFLGMRLQKLK
                LLQVSGCRQLKRLFSDQILAGTLPNLQEIKVVSCLRLEELFNFSSVPVDFCAESLLPK
                LTVIKLKYLPQLRSLCNDRVVLESLEHLEVESCESLKNLPFVPGNTGMINEQMAWEYM
                SRTLG"
exon            complement(9962. .12721)
                /gene="AT4g27220"
                /number=1
gene            13524. .14334
                /gene="AT4g27230"
gene            complement(join(13524. .13730,14146. .14334))
                /gene="AT4g27230"
CDS             complement(join(13524. .13730,14146. .14334))
                /gene="AT4g27230"
                /note="strong similarity to histone H2A.4, Triticum
                aestivum, PIR2:S53521
                Contains Histone H2A signature AA23-29
                contains EST gb:AI995882.1, T04821, T43418"
                /codon_start=1
                /product="histone H2A-like protein"
                /protein_id="CAA19717.1"
                /db_xref="GI:3269284"
                /db_xref="GOA:O81826"
                /db_xref="InterPro:IPR002119"
                /db_xref="InterPro:IPR007124"
                /db_xref="InterPro:IPR007125"
                /db_xref="UniProt/TrEMBL:O81826"
                /translation="MAGRGKQLGSGAAKKSTSRSSKAGLQFPVGRIARFLKAGKYAER
                VGAGAPVYLAAVLEYLAAEVLELAGNAARDNKKTRIVPRHIQLAVRNDEELSKLLGDV
                TIANGGVMPNIHNLLLPKKAGSSKPTEED"
exon            complement(13524. .13730)
                /gene="AT4g27230"
                /number=1

Alignment Scores:
Pred. No.:      1.66e+04        Length:         83371
Score:          42.00           Matches:        6
```

```
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     15             Gaps:          0

US-09-858-332G-28 (1-6) x ATM4I22 (1-83371)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      22858 CACAACCACAACCACAAC 22841

RESULT 512
AC091421/c
LOCUS       AC091421               84497 bp    DNA     linear   ROD 20-APR-2001
DEFINITION  Mus musculus chromosome 11 clone MGS1-124L2, complete sequence.
ACCESSION   AC091421
VERSION     AC091421.1  GI:13699327
KEYWORDS    HTG; HTGS_ACTIVEFIN.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 84497)
  AUTHORS   Resch,K., Platzer,M., Fuchs,S., Schilhabel,M., Jockusch,H.,
            Rosenthal,A. and Schmitt-John,T.
  TITLE     Chromosome 11 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 84497)
  AUTHORS   Platzer,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2001) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena, Thuringia 07745, Germany
FEATURES             Location/Qualifiers
     source          1. .84497
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="11"
                     /clone="MGS1-124L2"
                     /clone_lib="Genome Systems mouse ES cell BAC libraries I
                     and II"
     misc_feature    1. .4
                     /note="clone_end:SP6"
     misc_feature    84494. .84497
                     /note="clone_end:T7"
ORIGIN

Alignment Scores:
Pred. No.:              1.68e+04       Length:        84497
Score:                  42.00          Matches:       6
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     9              Gaps:          0

US-09-858-332G-28 (1-6) x AC091421 (1-84497)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      52036 CATAATCATAATCATAAT 52019

RESULT 513
AC106241_5
WPCOMMENT
  Sequence split into 6 fragments  LOCUS AC106241 Accession AC106241
     Fragment Name      Begin     End
     AC106241_0             1    110000
     AC106241_1        100001    210000
     AC106241_2        200001    310000
     AC106241_3        300001    410000
     AC106241_4        400001    510000
     AC106241_5        500001    584539

Continuation (6 of 6) of AC106241 from base 500001 (AC106241 Rattus norvegicus clone CH21

Alignment Scores:
Pred. No.:              1.68e+04       Length:        84539
Score:                  42.00          Matches:       6
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     14             Gaps:          0

US-09-858-332G-28 (1-6) x AC106241_5 (1-84539)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       2823 CACAACCACAACCACAAC 2840

RESULT 514
AC124117/c
LOCUS       AC124117               85693 bp    DNA     linear   HTG 09-JUN-2002
DEFINITION  Mus musculus clone RP23-9J7, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC124117
VERSION     AC124117.1  GI:21360239
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 85693)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-9J7
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 85693)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
            Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
            Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
            Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L26274
                Center clone name: 9_J_7
            --------------.
            * NOTE: This record contains 81 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
```

```
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*        1      972: contig of 972 bp in length
*      973     1072: gap of 100 bp
*     1073     2033: contig of 961 bp in length
*     2034     2133: gap of 100 bp
*     2134     3101: contig of 968 bp in length
*     3102     3201: gap of 100 bp
*     3202     4170: contig of 969 bp in length
*     4171     4270: gap of 100 bp
*     4271     5264: contig of 994 bp in length
*     5265     5364: gap of 100 bp
*     5365     6329: contig of 965 bp in length
*     6330     6429: gap of 100 bp
*     6430     7391: contig of 962 bp in length
*     7392     7491: gap of 100 bp
*     7492     8407: contig of 916 bp in length
*     8408     8507: gap of 100 bp
*     8508     9412: contig of 905 bp in length
*     9413     9512: gap of 100 bp
*     9513    10442: contig of 930 bp in length
*    10443    10542: gap of 100 bp
*    10543    11530: contig of 988 bp in length
*    11531    11630: gap of 100 bp
*    11631    12615: contig of 985 bp in length
*    12616    12715: gap of 100 bp
*    12716    13697: contig of 982 bp in length
*    13698    13797: gap of 100 bp
*    13798    14758: contig of 961 bp in length
*    14759    14858: gap of 100 bp
*    14859    15821: contig of 963 bp in length
*    15822    15921: gap of 100 bp
*    15922    16876: contig of 955 bp in length
*    16877    16976: gap of 100 bp
*    16977    17940: contig of 964 bp in length
*    17941    18040: gap of 100 bp
*    18041    19012: contig of 972 bp in length
*    19013    19112: gap of 100 bp
*    19113    19991: contig of 879 bp in length
*    19992    20091: gap of 100 bp
*    20092    21011: contig of 920 bp in length
*    21012    21111: gap of 100 bp
*    21112    22052: contig of 941 bp in length
*    22053    22152: gap of 100 bp
*    22153    23128: contig of 976 bp in length
*    23129    23228: gap of 100 bp
*    23229    24160: contig of 932 bp in length
*    24161    24260: gap of 100 bp
*    24261    25262: contig of 1002 bp in length
*    25263    25362: gap of 100 bp
*    25363    26350: contig of 988 bp in length
*    26351    26450: gap of 100 bp
*    26451    27430: contig of 980 bp in length
*    27431    27530: gap of 100 bp
*    27531    28492: contig of 962 bp in length
*    28493    28592: gap of 100 bp
*    28593    29520: contig of 928 bp in length
*    29521    29620: gap of 100 bp
*    29621    30563: contig of 943 bp in length
*    30564    30663: gap of 100 bp
*    30664    31619: contig of 956 bp in length
*    31620    31719: gap of 100 bp
*    31720    32629: contig of 910 bp in length
*    32630    32729: gap of 100 bp
*    32730    33684: contig of 955 bp in length
*    33685    33784: gap of 100 bp
*    33785    34747: contig of 963 bp in length
*    34748    34847: gap of 100 bp
*    34848    35858: contig of 1011 bp in length
*    35859    35958: gap of 100 bp
*    35959    36911: contig of 953 bp in length
*    36912    37011: gap of 100 bp
*    37012    37946: contig of 935 bp in length
*    37947    38046: gap of 100 bp
*    38047    39019: contig of 973 bp in length
*    39020    39119: gap of 100 bp
*    39120    40082: contig of 963 bp in length
*    40083    40182: gap of 100 bp
*    40183    41150: contig of 968 bp in length
*    41151    41250: gap of 100 bp
*    41251    42188: contig of 938 bp in length
*    42189    42288: gap of 100 bp
*    42289    43247: contig of 959 bp in length
*    43248    43347: gap of 100 bp
*    43348    44311: contig of 964 bp in length
*    44312    44411: gap of 100 bp
*    44412    45364: contig of 953 bp in length
*    45365    45464: gap of 100 bp
*    45465    46437: contig of 973 bp in length
*    46438    46537: gap of 100 bp
*    46538    47520: contig of 983 bp in length
*    47521    47620: gap of 100 bp
*    47621    48569: contig of 949 bp in length
*    48570    48669: gap of 100 bp
*    48670    49641: contig of 972 bp in length
*    49642    49741: gap of 100 bp
*    49742    50686: contig of 945 bp in length
*    50687    50786: gap of 100 bp
*    50787    51768: contig of 982 bp in length
*    51769    51868: gap of 100 bp
*    51869    52839: contig of 971 bp in length
*    52840    52939: gap of 100 bp
*    52940    53912: contig of 973 bp in length
*    53913    54012: gap of 100 bp
*    54013    54971: contig of 959 bp in length
*    54972    55071: gap of 100 bp
*    55072    56033: contig of 962 bp in length
*    56034    56133: gap of 100 bp
*    56134    57078: contig of 945 bp in length
*    57079    57178: gap of 100 bp
*    57179    58144: contig of 966 bp in length
*    58145    58244: gap of 100 bp
*    58245    59206: contig of 962 bp in length
*    59207    59306: gap of 100 bp
*    59307    60276: contig of 970 bp in length
*    60277    60376: gap of 100 bp
*    60377    61336: contig of 960 bp in length
*    61337    61436: gap of 100 bp
*    61437    62433: contig of 997 bp in length
*    62434    62533: gap of 100 bp
*    62534    63484: contig of 951 bp in length
*    63485    63584: gap of 100 bp
*    63585    64528: contig of 944 bp in length
*    64529    64628: gap of 100 bp
*    64629    65576: contig of 948 bp in length
*    65577    65676: gap of 100 bp
*    65677    66580: contig of 904 bp in length
*    66581    66680: gap of 100 bp
*    66681    67687: contig of 1007 bp in length
*    67688    67787: gap of 100 bp
*    67788    68738: contig of 951 bp in length
*    68739    68838: gap of 100 bp
*    68839    69794: contig of 956 bp in length
*    69795    69894: gap of 100 bp
*    69895    70794: contig of 900 bp in length
*    70795    70894: gap of 100 bp
*    70895    71831: contig of 937 bp in length
*    71832    71931: gap of 100 bp
*    71932    72885: contig of 954 bp in length
```

```
Alignment Scores:
Pred. No.:                1.7e+04        Length:          85693
Score:                    42.00          Matches:         6
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       14             Gaps:            0

US-09-858-332G-28 (1-6) x AC124117 (1-85693)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      61866 CACAACCACAACCACAAC 61849

RESULT 515
AC020021
LOCUS       AC020021               85699 bp    DNA     linear   HTG 03-JAN-2000
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC020021
VERSION     AC020021.1  GI:6664876
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 85699)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10211768 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .85699
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
ORIGIN

Alignment Scores:
Pred. No.:                1.7e+04        Length:          85699
Score:                    42.00          Matches:         6
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       14             Gaps:            0

US-09-858-332G-28 (1-6) x AC020021 (1-85699)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      34846 CACAATCACAACCACAAT 34863

RESULT 516
AP002047
LOCUS       AP002047               86536 bp    DNA     linear   PLN 14-FEB-2004
DEFINITION  Arabidopsis thaliana genomic DNA, chromosome 3, P1 clone: MQC3.
ACCESSION   AP002047
VERSION     AP002047.1  GI:8051655
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Kaneko,T., Katoh,T., Sato,S., Nakamura,A., Asamizu,E. and Tabata,S.
  TITLE     Structural analysis of Arabidopsis thaliana chromosome 3. II.
            Sequence features of the 4,251,695 bp regions covered by 90 P1, TAC
            and BAC clones
  JOURNAL   DNA Res. 7 (3), 217-221 (2000)
  PUBMED    10907853
REFERENCE   2  (bases 1 to 86536)
  AUTHORS   Kaneko,T., Kato,T., Sato,S., Nakamura,Y., Asamizu,E. and Tabata,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2000) Yasukazu Nakamura, Kazusa DNA Research
            Institute, Department of Plant Gene Research; 1532-3, Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
            Tel:81-438-52-3935, Fax:81-438-52-3934)
COMMENT     Address for correspondence: kaos@kazusa.or.jp
            For the latest information on annotation of this clone, please see
            http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MQC3
            Genes with similarity to proteins in the databases are described in
            'product' or 'note' qualifiers.  Genes that have no significant
            protein similarity are described as 'unknown protein'.
            The software programs used to predict genes include: Grail
            (Informatics Group, Oak Ridge National Laboratory,
            http://compbio.ornl.gov/Grail-1.3/),
            GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
            NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
            Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
            SplicePredictor (Volker Brendel, Stanford University,
            http://gremlin1.zool.iastate.edu/cgi-bin/sp.cgi).
            Genes encoding tRNAs are predicted by tRNAscan-SE
            (Sean Eddy, Washington University School of Medicine, St. Louis,
            http://genome.wustl.edu/eddy/tRNAscan-SE/).
            This sequence may not be the entire insert of this clone. It may be
            shorter because we remove overlaps between neighboring submissions.
            The 5' clone is T23B7 and the 3' clone is T16H11.
FEATURES             Location/Qualifiers
     source          1. .86536
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /chromosome="3"
                     /clone="MQC3"
                     /clone_lib="Mitsui P1"
                     /ecotype="Columbia"
     CDS             complement(join(844. .1770,1859. .1910,1999. .2072,
                     2167. .2216,2289. .2370,2491. .2540,2638. .2680,2782. .2862,
                     2968. .3013,3123. .3169,3247. .3312,3724. .3790,4091. .4221))
                     /note="unnamed protein product; contains similarity to
                     protein kinase
                     gene_id:MQC3.2"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAB03128.1"
                     /db_xref="GI:15795140"
                     /translation="MQMEANDCQEEHKFTLDNYHVVEQVRRGKSSSDFVVLHDIEDKK
                     YAMKKICLAKHTDKLKQTALQEMKLLSSLKNPYIVHYEDSWIDNDNNACIFTAYYEGG
                     NMANAIKKARGKLFPEERIFKWLAQLLLAVNYLHSNRVVHMDLTCSNIFLPKDDHVQL
                     GNYGLAKLINPEKPVSMVSGISNSMCPEVLEDQPYGYKSDIWSLGCCMYEITAHQPAF
                     KAPDMAGLINKINRSLMSPLPIVYSSTLKQMIKLMLRKKPEYRPTACELLRNPSLQPY
                     LLQCQNLSPIYLPVFPIKPVNSPKDKARRNSLPGKFGKERVSREKSEVSRSLENLYPF
                     WTNTETGSSSSSQPASSTNGAEDKLETKRIDPSCDTLKISEFTSQKSDESLIDPDIAV
                     YSTETPAEENALPKETENIFSEESQLRDVDVGVVSAQEVACSPPRAIEEAETQEALPK
                     PKEQITVPISSVAHSSTEVAAAKDHLSGSLEGDKAKMVKLTASEMSSVLSKLTKLGPP
                     QSKERADALECLLEKCAGLVKQEKYEELAGLLTPFGEDGVSARDTAIWFAKTLLSSDK
                     LNQGT"
     CDS             complement(join(5028. .5108,5218. .5263,5441. .5554,
                     5664. .5782,5869. .5943,6031. .6093,6202. .6288,6363. .6464,
                     6559. .6596,6680. .6782,6876. .6998,7092. .7218,7488. .7627))
                     /note="gene_id:MQC3.3"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="serine carboxypeptidase"
                     /protein_id="BAB03129.1"
                     /db_xref="GI:15795141"
                     /translation="MGKECYYLSWILKFHLLLVLIQLVDSGSTIRFLPGFQGPLPFEL
                     ETGISDLFELQTTTKNKNAGPIAFKAEEYNGSIPSLVSTTYAWTKVASILYLDQPVGT
                     GFSYSRNPLADIPSDTGVAKPVNEFLHKWLDKHPEFLSNPLYVAGNSYSGIVIPTIVQ
```

EISNGNHLDSKPQINLQGFVLGNPATDTDIDLNSRIPFAHGKALISDEHYESLKRSCQ
GNYISVNPRNTKCLKLLEDFKKCVSGISEEYILKPDCMWLYSCMANLHSLSEYWANEK
SVRKALLVNEGTVRKWIRCNTEIAYNKDIRSSVPYHKYISIEGYRSLVFSGDHDMLVP
FLGTQAWIRSLNYSIVDDWRPWMVQNQVAGYTRTYANKMTFATVKGGGHTSEYKPVET
YIMIKRWLSGQPL"

CDS join(8514. .8737,8837. .9032,9127. .9213,9428. .9546,
9651. .9763,9837. .9976)
/note="unnamed protein product; gb|AAF48080.1
gene_id:MQC3.4
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB03130.1"
/db_xref="GI:15795142"
/translation="MELEFKCSDVGVWKEALSSYDSRIESLNKPELVSLDQFYRIKLP
CLLHDRDPNPYLTTSELSQLMKWKLSRGKWRPRLLDFVSSLDDSVVKSASEKAFKSLP
DISKAVKELTVLKGVGAATASAVLAAYAPDIAPFMSDEAMEVALGNSKDYSLKQYLLF
ATKLQDKAKELKLKGEWDGPSDIERALWSCTVRAKSQPEKSSSGKKRKRKCSRMFYQN
AEYLAESYNLKGKNGYFTFSDIFQINIEGYEGRVIGGEKQNRGYVVSHLTKLILACII
ADRLLLVTEFGHLSIP"

CDS complement(join(10264. .10417,10502. .10615,10701. .10819,
10897. .11115,11201. .11287,11367. .11468,11557. .11594,
11680. .11782,11868. .11990,12080. .12156,12237. .12423))
/note="gene_id:MQC3.5"
/codon_start=1
/evidence=not_experimental
/product="serine carboxypeptidase"
/protein_id="BAB03131.1"
/db_xref="GI:15795143"
/translation="MTLCSLSFGVKSLRYIGVGEEDEDQMFYYFIKSESNPKTDPLLL
WLSGGPGCSSFTGLIYENGPLGFKVEAYNGSIPTLVSTTYSWTKVANIIYLDQPVGTG
FSYSRNPLADIPSDTGSAKRVDEFLRKWLTKHPEYFSNPFYAGGNSYSGKMVPVIVQE
ISNGNCIYGKPQIRLQGYVLGSPVTDYDLDRNSRIQFAHGMALISNELYESMKRTCGG
NYIFVDPLNTECLELIKDYDNCVSGIYENLILVPKCDLTSPDCHVSILVDFALATKLN
HKRRIFTKLLSQSYRSMLSDYWANNESVRRALKVVEGTTGRWERCKWTLQNNKDIKSS
IPYHKKNSIQGYRSLIFSGDHDMLTPYVGTQDWIRSLNYSIIDKWRPWMILDQVAGYT
TTYANKMTFATVKVTPISTFSIDKNAKLFYITLLNRFVGLICRVAGTR"

CDS complement(join(13183. .13327,13431. .13544,13640. .13758,
13831. .13902,14002. .14070,14156. .14242,14332. .14433,
14521. .14558,14639. .14741,14834. .14956,15042. .15118,
15201. .15346,15420. .15550))
/note="gene_id:MQC3.6"
/codon_start=1
/evidence=not_experimental
/product="serine carboxypeptidase"
/protein_id="BAB03132.1"
/db_xref="GI:15795144"
/translation="MGSWIPKLLLLQLVLLLTKHADSSSIIKYLPGFEGPLPFELETG
YIGVGDEDEDQMFYYFIKSESNPEEDPLLVWLSGGPGCSSFTGLVYENGPLGFKVEAY
NGSIPTLVSTTYSWTKVANIIYLDQPVGAGFSYSRNPFADRPSDTGSAKLVNEFVRKW
LAKHPDYFSNPFYVTGNSYSGKVIPAIVQEISNGNYICCKPQINLQGYVIGNPVAYYD
HDKDSRIPFAHGVALISDELFESLKRSCGGSYSIVDPLNTECLKLIKDYHKCVSGIYQ
ELILKPKCETTSPDCYTYRYLLSIYWANNEIVRRALKVVEGSKGKWERCDLSVRSNQD
IKSSIPYHMNNSIKGYRSLVISGDHDMTIPFLGTQAWIRSLNYSITEKWRPWMILDQV
AGYTKTYANKMTLATVKVCSFIHILIFFVNIKIESCFTSVVIGLWVWLL"

CDS complement(join(16107. .16187,16282. .16327,16418. .16531,
16626. .16744,16819. .16890,16999. .17067,17157. .17243,
17345. .17446,17541. .17578,17660. .17762,17846. .17968,
18059. .18135,18222. .18367,18456. .18589))
/note="gene_id:MQC3.7"
/codon_start=1
/evidence=not_experimental
/product="serine carboxypeptidase"
/protein_id="BAB03133.1"
/db_xref="GI:15795145"
/translation="MASWIFKLLLLLQCVLVLIQHADSSSIIRYLPGFEGPLPFELET
GYIGVGQKEEDQLFYYFIKSENNPEEDPLLVWLTGGPGCSSFSGLVYENGPLAFKVET
YNGSVPTLVSTTYSWTKVANIIYLDQPVGTGFSYSRNPFADIPSDTGSVKRVNEFVRK
WLAKHPEYFSNPFYVTGNSYSGKVIPAIVQEISNGNYICCKPQINLQGYVIGNPVAYY
DHDKDFRIPFAHGVALISDELFESLKASCGGSYSVVDPLNTECLKLIEDYDKCVSGIY
EELILKSKCEHTSPDCYTYRYLLSEYWADNETVRRALKVVKGSKGTWERCDYRVLSNQ
DIKSSIPFHINNSIRGYRSLVISGDHDMTIPFLGTQAWIRSLNYSITEKWRPWMILDQ
VAGYTKTYANKMTLATVKGGGHTLEYKPEENSVLFKRWISGQPL"

CDS join(20244. .20363,20455. .20514,20639. .20716,21050. .21105,
21392. .21662,21737. .21982,22108. .22254)
/note="gene_id:MQC3.8"
/codon_start=1
/evidence=not_experimental
/product="leucine zipper transcription factor HBP-1b"
/protein_id="BAB03134.1"
/db_xref="GI:15795146"
/translation="MHSLNETVIPDVDYMQSDRGHMHAAASDSSDRSKDKLDQKTLRR
LAQNREAARKSRLRKKAYVQQLENSRLKLTQLEQELQRARQQGVFISSSGDQAHSTGG
NGGALAFDAEHSRWLEEKNRQMNELRSALNAHAGDTELRIIVDGVMAHYEELFRIKSN
AAKNDVFHLLSGMWKTPAERCFLWLGGFRSSELLKLLANQLEPMTERQVMGINSLQQT
SQQAEDALSQGMESLQQSLADTLSSGTLGSSSSDNVASYMGQMAMAMGQLGTLEGFIR
QADNLRLQTLQQMLRVLTTRQSARALLAIHDYSSRLRALSSLWLARPRE"

CDS complement(join(22923. .23105,23790. .24008))
/note="unnamed protein product; gene_id:MQC3.9
similar to unknown protein
sp|Q02366"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB03135.1"
/db_xref="GI:15795147"
/translation="MAAPFALRKIGVPPNSANLTEARRRVFDFFRAACRSIPTIMDIY
NLQDVVAPSQLRYAISAQIRNNAHITDPKVIDLLIFKGMEELTDIVDHAKQRHHIIGQ
YVVGEGLVQNTGNKDQGKTDFLKNFYTSNYF"

CDS join(24313. .24630,24711. .25178,25294. .25391,25624. .25702,
25784. .26026,26114. .26281,26359. .26793)
/note="gene_id:MQC3.10"
/codon_start=1
/evidence=not_experimental
/product="protein arginine N-methyltransferase 3-like
protein"
/protein_id="BAB03136.1"
/db_xref="GI:15795148"
/translation="MAATMVKHEILNYSEDEEENYSDEGDWGDWKADDNGIEGGEEEE
EDDGDDSESDFLCLFCDSHFVSCDLLFEHCRLSHGFDFHGVRKELKLDFYSSFKLINY
IRSQVAENMCFSWKIEADDYKDVKFPWDEEKYLKPFWQEDSLLYSFADDEEDEEVTFD
REEVMEELQKLGDLSIDVEALGESSMSNSDKCNINGSKDVTSLSNCNGLKQSSADDLI
VNGKDAEPKVCDGRLVNRNIRKVNENYFGSYSSFGIHREMLSDKVRTEAYRDALLKNP
TLLNGSVVMDVGCGTGILSLFAAKAGASRVVAVEASEKMAKVATKQIAKDNKVFNDNE
HNGVLEVAHSMVEELDKSIQIQPHSVDVLVSEWMGYCLLYESMLSSVLYARDRWLKPG

Alignment Scores:

Pred. No.:	1.72e+04	Length:	86536
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	15	Gaps:	0

US-09-858-332G-28 (1-6) x AP002047 (1-86536)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      73802 CACAACCACAACCACAAC 73819
```

RESULT 517
AC006587/c
LOCUS AC006587 86579 bp DNA linear PLN 11-MAR-2002
DEFINITION Arabidopsis thaliana chromosome 2 clone T17D12 map mi54, complete sequence.
ACCESSION AC006587
VERSION AC006587.5 GI:20197880
KEYWORDS HTG.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 86579)
AUTHORS Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Shen,M., VanAken,S.E., Barnstead,M.E., Mason,T.M., Bowman,C.L., Ronning,C.M., Benito,M.-I., Carrera,A.J., Creasy,T.H., Buell,C.R., Town,C.D.,

```
            Nierman,W.C., Fraser,C.M. and Venter,J.C.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 86579)
  AUTHORS   Lin,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 86579)
  AUTHORS   Town,C.D. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
COMMENT     On Apr 18, 2002 this sequence version replaced gi:6598612.
            On Oct 8, 1997 this sequence version replaced gi:2443866.
             We have determined that YAC YUP8H12 is chimaeric, and is comprised
             of two distinct genomic EcoRI fragments from chromosome I. This
             submission contains the sequence from the EcoRI site at position 1

             (right end) to position 181918 of our previous Phase II
             submission YUP8H12 accession number AC000098. This fragment maps
            at
             the bottom of the right arm of the chromosome between the ATHATPAS
             and m532 markers.
FEATURES             Location/Qualifiers
     source          1. .86579
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /chromosome="2"
                     /map="mi54"
                     /clone="T17D12"
                     /ecotype="Columbia"
     misc_feature    complement(1. .5360)
                     /note="overlap with BAC clone T1B3 (AC006283:1. .5360)."
     gene            1557. .3671
                     /gene="At2g28460"
                     /note="synonym: T17D12.2; predicted by genscan"
     mRNA            <1557. .>3671
                     /gene="At2g28460"
     CDS             1557. .3671
                     /gene="At2g28460"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAD21481.1"
                     /db_xref="GI:4510394"
                     /translation="MSPAKQIRFDLDWVGEKRMNSFITKVMSLVSSMDLDLQPKPESE
                     FMSLTTQIFSLLHSMDSDSMPKPLSNLISLLSQGNFDHNTDFRLFFRQTMALEPEPIF
                     MSLIYQIFSLVISMNTKRDKLISLCPQACVVLGNNGNFEVIVKKQEGQAHKGKAIPWS
                     RDSKWECLPFNWKTYWSSGDDVTHFRCRNCDGDNHKEYEKAPVEIKHSLHRKHSLQLV
                     LLDESSYTRVCYCCDEDLKRMFYYCRTCDFGLNFVCAKKEAILYIDQPKWHEHTLALF
                     LRKTSLTCNACGLSHSSCPLYMCPPCDFVIHKSCISLPRLIRISRHFHRIAYTPSFDE
                     GDWSCSVCRKKIDNDYGGYVCTKGCSYAAHSKCATQSNVWDGIELEGEPEDIEEEVLP
                     PFLEISDGIIQHFSHQQHHMKLDENTGRDYDENKECEACIRPIYFGNFYSCLECDFIL
                     HEECANLSRKIHHPIHPHLLNLIGGFDGVINYYNDKCSACIGLCKGGFFYECGKQGCK
                     FMLHVQCATTSEPLVHESHRHPLFLTSKPGEKIRCSVCKDSEETFNCIECDFALCFYC
                     AILPQKVRYKHDKHTLTLSYGKETGTSWCEVCEAKINLKGRFYMCNEYCCVTLHIKCT
                     LGDLHMRPGSTWITFGNKVDVLHNSNRMSRSICSYCVKHCPYKIVYMCSGLILCSQGC
                     VVRSILQKRGRYRELEFLSRYA"
     gene            complement(3707. .8000)
                     /gene="At2g28470"
                     /note="synonym: T17D12.3"
     mRNA            complement(join(<3707. .4008,4100. .4445,4517. .4728,
                     4813. .4914,4983. .5096,5178. .5278,5387. .5496,5604. .5758,
                     5858. .6049,6126. .6244,6326. .6413,6517. .6622,6693. .6781,
                     6888. .7031,7120. .7212,7295. .7361,7464. .7576,7667. .7762,
                     7836. .>8000))
                     /gene="At2g28470"
     CDS             complement(join(3901. .4008,4100. .4445,4517. .4728,
                     4813. .4914,4983. .5096,5178. .5278,5387. .5496,5604. .5758,
                     5858. .6049,6126. .6244,6326. .6413,6517. .6622,6693. .6781,
                     6888. .7031,7120. .7212,7295. .7361,7464. .7576,7667. .7762,
                     7836. .8000))
                     /gene="At2g28470"
                     /codon_start=1
                     /product="putative beta-galactosidase"
                     /protein_id="AAD21482.1"
                     /db_xref="GI:4510395"
                     /translation="MVKVRKMEMILLLILVIVVAATAANVTYDHRALVIDGKRKVLIS
                     GSIHYPRSTPEMWPELIQKSKDGGLDVIETYVFWSGHEPEKNKYNFEGRYDLVKFVKL
                     AAKAGLYVHLRIGPYVCAEWNYGGFPVWLHFVPGIKFRTDNEPFKEEMQRFTTKIVDL
                     MKQEKLYASQGGPIILSQIENEYGNIDSAYGAAAKSYIKWSASMALSLDTGVPWNMCQ
                     QTDAPDPMINTCNGFYCDQFTPNSNNKPKMWTENWSGWFLGFGDPSPYRPVEDLAFAV
                     ARFYQRGGTFQNYYMYHGGTNFDRTSGGPLISTSYDYDAPIDEYGLLRQPKWGHLRDL
                     HKAIKLCEDALIATDPTITSLGSNLEAAVYKTESGSCAAFLANVDTKSDATVTFNGKS
                     YNLPAWSVSILPDCKNVAFNTAKVKFNSISKTPDGGSSAELGSQWSYIKEPIGISKAD
                     AFLKPGLLEQINTTADKSDYLWYSLRTDIKGDETFLDEGSKAVLHIESLGQVVYAFIN
                     GKLAGSGHGKQKISLDIPINLVTGTNTIDLLSVTVGLANYGAFFDLVGAGITGPVTLK
                     SAKGGSSIDLASQQWTYQVGLKGEDTGLATVDSSEWVSKSPLPTKQPLIWYKTTFDAP
                     SGSEPVAIDFTGTGKGIAWVNGQSIGRYWPTSIAGNGGCTESCDYRGSYRANKCLKNC
                     GKPSQTLYHVPRSWLKPSGNILVLFEEMGGDPTQISFATKQTGSNLCLTVSQSHPPPV
                     DTWTSDSKISNRNRTRPVLSLKCPISTQVIFSIKFASFGTPKGTCGSFTQGHCNSSRS
                     LSLVQKACIGLRSCNVEVSTRVFGEPCRGVVKSLAVEASCS"
repeat_region        7581. .7654
                     /rpt_family="(CAAAAA)n"
gene                 complement(11496. .12885)
                     /gene="At2g28480"
                     /note="synonym: T17D12.4; predicted by genscan and
                     genefinder"
mRNA                 complement(join(<11496. .11810,11908. .12309,12399. .12720,
                     12806. .>12885))
                     /gene="At2g28480"
CDS                  complement(join(11496. .11810,11908. .12309,12399. .12720,
                     12806. .12885))
                     /gene="At2g28480"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAD21483.1"
                     /db_xref="GI:4510396"
                     /translation="MAIAFARGLRKASSFLRSSYPLLISSRLLPISQEFYDSTLTFVL
                     PTRPICRNFSHGYGTVNLVISEGKPKFETRELDPPKKYKWLTKKRLKLKRKKEREERN
                     AANRKDPRRLTVKGKKKKFVNAEERIKYKLEKAKIKEALLIERLKRYEVAKVQGPEVR
                     PHEITGEERFYLKKMGQKRSNYVPIGRRGVFGGVILNMHLHWKKHETVKVICNNSKPG
                     QVQQYAEELAKLSGGVPVNIIGDDTIIFYRGKGYVQPQVMSPIDTLSKKRAYEKSKYE
                     QSLESVRHFIAIAEKELELYYRHVALYDDPNNRNPLSILDDSPSESRQHHENEFFLSC
                     SDTDSSSEDEDEDEELCELDNGSSSSSSAKEELSESEN"
gene                 complement(13451. .15837)
                     /gene="At2g28490"
                     /note="synonym: T17D12.5"
mRNA                 complement(join(<13451. .14085,14335. .14757,14987. .15061,
                     15134. .15315,15402. .>15837))
                     /gene="At2g28490"
CDS                  complement(join(13666. .14085,14335. .14757,14987. .15061,
                     15134. .15315,15402. .15837))
                     /gene="At2g28490"
                     /codon_start=1
                     /product="putative seed storage protein (vicilin-like)"
                     /protein_id="AAD21484.1"
                     /db_xref="GI:4510397"
                     /translation="MEKNKRAIGFLLLVVLINGVMMTRSNGYEGEEEWGGAGGGEWGG
                     AEGGGAWGGGGGGGGAWGGEGEGGGEWGGGGEGGGGGRRGWFMMRESRQVIKSEGGEM
                     RVVLSPRGRIIEKPMHIGFLTMEPKTLFVPQYLDSSLLIFIRQGEATLGVICKDEPGE
                     RKLKAGDIYWIPAGSVFYLHNTGLGQRLHVICSIDPTQSLGFETFQPFYIGGGPSSVL
                     AGFDPHTLTSAFNVSLPELQQMMMSQFRGPIVYVTEGPQPQPQSTVWTQFLGLRGEEK
                     HKQLKKLLETKQGSPQDQQYSSGWSWRNIVRSILDLTEEKNKGSGSSECEDSYNIYDK
                     KDKPSFDNKYGWSIALDYDDYKPLKHSGIGVYLVNLTAGAMMAPHMNPTATEYGIVLA
                     GSGEIQVVFPNGTSAMNTRVSVGDVFWIPRYFAFCQIASRTGPFEFVGFTTSAHKNRP
                     QFLVGSNSLLRTLNLTSLSIAFGVDEETMRRFIEAQREAVILPTPAAAPPHVGEMVSD
                     RFV"
repeat_region        complement(14887. .14911)
                     /rpt_family="AT_rich"
repeat_region        complement(15586. .15751)
                     /rpt_family="(GGA)n"
repeat_region        complement(16265. .16336)
                     /rpt_family="AT_rich"
repeat_region        complement(17104. .17208)
                     /rpt_family="AT_rich"
```

```
     repeat_region   complement(17242. .17274)
                     /rpt_family="AT_rich"
     repeat_region   19368. .19403
                     /rpt_family="AT_rich"
     repeat_region   complement(21016. .21040)
                     /rpt_family="(A)n"
     gene            complement(21457. .22955)
                     /gene="At2g28500"
                     /note="synonym: T17D12.6; predicted by genscan and
                     genefinder"
     mRNA            complement(join(<21457. .21852,22662. .>22955))
                     /gene="At2g28500"
     CDS             complement(join(21457. .21852,22662. .22955))
                     /gene="At2g28500"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAD21485.1"
                     /db_xref="GI:4510398"
                     /translation="MEINGGVATPTASAVAKVTETTTPVNSPSPTSSPPPPPSPQQPP
                     QPPVVLSPCAACKILRRRCADKCVLAPYFPPTDPAKFTIAHRVFGASNIIKFLQELPE
                     SQRTDAVNSMVYEAGARMRDPVYGCAGAIYHLQRQVSELQAQLAKTQVELVGMQLQRS
                     SLLELIYNMEQTKLSVQEQGQQKMSFESSFESGDEFISSPDEESNDLGFLEDNNNNNN
                     SSMSWWDPLWT"
     repeat_region   21856. .21877
                     /rpt_family="AT_rich"
     repeat_region   complement(22357. .22388)
                     /rpt_family="AT_rich"
     repeat_region   25281. .25358
                     /rpt_family="(TAAAA)n"
     repeat_region   complement(27274. .27295)
                     /rpt_family="AT_rich"
     repeat_region   complement(27814. .27854)
                     /rpt_family="(CATA)n"
     repeat_region   complement(28297. .28390)
                     /rpt_family="(TA)n"
     repeat_region   complement(30169. .30204)
                     /rpt_family="AT_rich"
     repeat_region   complement(30601. .30621)
                     /rpt_family="AT_rich"

Alignment Scores:
Pred. No.:              1.72e+04        Length:         86579
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x AC006587 (1-86579)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      14684 CACAACCACAACCACAAT 14667

RESULT 518
AC151664/c
LOCUS       AC151664               86869 bp    DNA     linear   HTG 19-MAY-2005
DEFINITION  Medicago truncatula clone mth2-139m20, WORKING DRAFT SEQUENCE, 11
            unordered pieces.
ACCESSION   AC151664
VERSION     AC151664.17  GI:66275404
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 86869)
  AUTHORS   Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
            and Roe,B.A.
  TITLE     Medicago truncatula BAC Clone mth2-139m20
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 86869)
  AUTHORS   Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
            and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-2004) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 86869)
  AUTHORS   Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
            and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAY-2005) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On May 19, 2005 this sequence version replaced gi:63003962.
            -------------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ----------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2187: contig of 2187 bp in length
            *     2188     2287: gap of unknown length
            *     2288     6238: contig of 3951 bp in length
            *     6239     6338: gap of unknown length
            *     6339     9805: contig of 3467 bp in length
            *     9806     9905: gap of unknown length
            *     9906    13302: contig of 3397 bp in length
            *    13303    13402: gap of unknown length
            *    13403    16872: contig of 3470 bp in length
            *    16873    16972: gap of unknown length
            *    16973    25035: contig of 8063 bp in length
            *    25036    25135: gap of unknown length
            *    25136    31808: contig of 6673 bp in length
            *    31809    31908: gap of unknown length
            *    31909    42068: contig of 10160 bp in length
            *    42069    42168: gap of unknown length
            *    42169    52548: contig of 10380 bp in length
            *    52549    52648: gap of unknown length
            *    52649    65724: contig of 13076 bp in length
            *    65725    65824: gap of unknown length
            *    65825    86869: contig of 21045 bp in length.
FEATURES            Location/Qualifiers
     source         1. .86869
                    /organism="Medicago truncatula"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:3880"
                    /clone="mth2-139m20"
                    /clone_lib="Medicago truncatula BAC library H2"
     gap            2188. .2287
                    /estimated_length=unknown
     gap            6239. .6338
                    /estimated_length=unknown
     gap            9806. .9905
                    /estimated_length=unknown
     gap            13303. .13402
                    /estimated_length=unknown
     gap            16873. .16972
                    /estimated_length=unknown
     gap            25036. .25135
                    /estimated_length=unknown
     gap            31809. .31908
                    /estimated_length=unknown
     gap            42069. .42168
                    /estimated_length=unknown
     gap            52549. .52648
```

```
                     /estimated_length=unknown
     gap             65725. .65824
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:                1.72e+04        Length:          86869
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       14              Gaps:            0

US-09-858-332G-28 (1-6) x AC151664 (1-86869)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1981 CATAATCATAATCATAAT 1964

RESULT 519
AP007492/c
LOCUS       AP007492               87230 bp    DNA     linear   HTG 28-DEC-2004
DEFINITION  Lotus corniculatus var. japonicus chromosome 1 clone LjT14F08, ***
            SEQUENCING IN PROGRESS ***, 12 unordered pieces.
ACCESSION   AP007492
VERSION     AP007492.1  GI:56805806
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1
  AUTHORS   Kaneko,T., Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Structural Analysis of a Lotus japonicus Genome. XI. Sequence
            Features and Mapping of Nine hundred twenty-one TAC Clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 87230)
  AUTHORS   Sato,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-2004) Shusei Sato, Kazusa DNA Research Institute,
            Department of Plant Gene Research; 2-6-7 Kazusa-kamatari, Kisarazu,
            Chiba, 292-0818, Japan (E-mail:ssato@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/, Tel:81-438-52-3935(ex.2337),
            Fax:81-438-52-3934)
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    15414: contig of 15414 bp in length
            *    15415    15514: gap of unknown length
            *    15515    16477: contig of 963 bp in length
            *    16478    16577: gap of unknown length
            *    16578    17714: contig of 1137 bp in length
            *    17715    17814: gap of unknown length
            *    17815    19353: contig of 1539 bp in length
            *    19354    19453: gap of unknown length
            *    19454    22164: contig of 2711 bp in length
            *    22165    22264: gap of unknown length
            *    22265    24958: contig of 2694 bp in length
            *    24959    25058: gap of unknown length
            *    25059    29566: contig of 4508 bp in length
            *    29567    29666: gap of unknown length
            *    29667    36383: contig of 6717 bp in length
            *    36384    36483: gap of unknown length
            *    36484    44587: contig of 8104 bp in length
            *    44588    44687: gap of unknown length
            *    44688    54174: contig of 9487 bp in length
            *    54175    54274: gap of unknown length
            *    54275    60896: contig of 6622 bp in length
            *    60897    60996: gap of unknown length
            *    60997    87230: contig of 26234 bp in length.
FEATURES             Location/Qualifiers
     source          1. .87230
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="genomic DNA"
                     /variety="japonicus"
                     /db_xref="taxon:34305"
                     /chromosome="1"
                     /clone="LjT14F08"
                     /clone_lib="LjT library"
                     /note="TAC clone:TM0653, synonym:Lotus japonicus"
     gap             15415. .15514
                     /estimated_length=unknown
     gap             16478. .16577
                     /estimated_length=unknown
     gap             17715. .17814
                     /estimated_length=unknown
     gap             19354. .19453
                     /estimated_length=unknown
     gap             22165. .22264
                     /estimated_length=unknown
     gap             24959. .25058
                     /estimated_length=unknown
     gap             29567. .29666
                     /estimated_length=unknown
     gap             36384. .36483
                     /estimated_length=unknown
     gap             44588. .44687
                     /estimated_length=unknown
     gap             54175. .54274
                     /estimated_length=unknown
     gap             60897. .60996
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:                1.73e+04        Length:          87230
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       14              Gaps:            0

US-09-858-332G-28 (1-6) x AP007492 (1-87230)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      52883 CACAATCACAATCACAAT 52866

RESULT 520
AC014497
LOCUS       AC014497               87767 bp    DNA     linear   HTG 16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC014497
VERSION     AC014497.1  GI:6436838
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 87767)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10210380 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
```

```
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .87767
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
ORIGIN

Alignment Scores:
Pred. No.:              1.74e+04        Length:         87767
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AC014497 (1-87767)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      17155 CACAATCACAATCACAAT 17172

RESULT 521
AP007309/c
LOCUS       AP007309               87776 bp    DNA     linear   HTG 28-DEC-2004
DEFINITION  Lotus corniculatus var. japonicus chromosome 1 clone LjT48M16, ***
            SEQUENCING IN PROGRESS ***, 7 unordered pieces.
ACCESSION   AP007309
VERSION     AP007309.1  GI:56805625
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1
  AUTHORS   Kaneko,T., Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Structural Analysis of a Lotus japonicus Genome. XI. Sequence
            Features and Mapping of Nine hundred twenty-one TAC Clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 87776)
  AUTHORS   Sato,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-2004) Shusei Sato, Kazusa DNA Research Institute,
            Department of Plant Gene Research; 2-6-7 Kazusa-kamatari, Kisarazu,
            Chiba, 292-0818, Japan (E-mail:ssato@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/, Tel:81-438-52-3935(ex.2337),
            Fax:81-438-52-3934)
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     5928: contig of 5928 bp in length
            *     5929     6028: gap of unknown length
            *     6029     7053: contig of 1025 bp in length
            *     7054     7153: gap of unknown length
            *     7154    11485: contig of 4332 bp in length
            *    11486    11585: gap of unknown length
            *    11586    23953: contig of 12368 bp in length
            *    23954    24053: gap of unknown length
            *    24054    45548: contig of 21495 bp in length
            *    45549    45648: gap of unknown length
            *    45649    79218: contig of 33570 bp in length
            *    79219    79318: gap of unknown length
            *    79319    87776: contig of 8458 bp in length.
FEATURES             Location/Qualifiers
     source          1. .87776
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="genomic DNA"
                     /variety="japonicus"
                     /db_xref="taxon:34305"
                     /chromosome="1"
                     /clone="LjT48M16"
                     /clone_lib="LjT library"
                     /note="TAC clone:TM1524, synonym:Lotus japonicus"
     gap             5929. .6028
                     /estimated_length=unknown
     gap             7054. .7153
                     /estimated_length=unknown
     gap             11486. .11585
                     /estimated_length=unknown
     gap             23954. .24053
                     /estimated_length=unknown
     gap             45549. .45648
                     /estimated_length=unknown
     gap             79219. .79318
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              1.74e+04        Length:         87776
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AP007309 (1-87776)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       8952 CACAACCACAACCACAAC 8935

RESULT 522
AP002957
LOCUS       AP002957               87834 bp    DNA     linear   PRI 15-MAR-2003
DEFINITION  Homo sapiens genomic DNA, chromosome 11q, clone:CTD-2337I7,
            complete sequences.
ACCESSION   AP002957
VERSION     AP002957.2  GI:13810522
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA
  JOURNAL   Published Only in Database (2000)
REFERENCE   2  (bases 1 to 87834)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-NOV-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     On Apr 26, 2001 this sequence version replaced gi:11526584.
FEATURES             Location/Qualifiers
     source          1. .87834
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="CTD-2337I7"
ORIGIN
```

```
Alignment Scores:
Pred. No.:                1.74e+04        Length:          87834
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       8               Gaps:            0

US-09-858-332G-28 (1-6) x AP002957 (1-87834)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      81549 CATAATCATAATCATAAT 81566

RESULT 523
AC116924/c
LOCUS       AC116924               88549 bp    DNA     linear   INV 21-FEB-2004
DEFINITION  Dictyostelium discoideum chromosome 2 map 6357117-6445670 strain
            AX4, complete sequence.
ACCESSION   AC116924
VERSION     AC116924.3  GI:42733690
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 88549)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 88549)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 88549)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 88549)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   5  (bases 1 to 88549)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2004) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On Feb 21, 2004 this sequence version replaced gi:28829074.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES              Location/Qualifiers
     source           1. .88549
                      /organism="Dictyostelium discoideum"
                      /mol_type="genomic DNA"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /chromosome="2"
                      /map="6357117-6445670"
     CDS              1864. .2241
                      /note="GeneID exon scores (in order of location ranges):
                      27.07 - GSCJ_ID dd_01259"
                      /codon_start=1
                      /product="hypothetical protein"
                      /protein_id="AAO51640.2"
                      /db_xref="GI:42733691"
                      /translation="MDQFKVPPPTYAPVAGQTIYGAPSYNNYYMRTPQVSTAPMIYPT
                      PMMAPPIMTPPIMTQPIMTPPMMYPPIIPSQPFMGPSMVSPIMSPPMIPSQPFMGPSM
                      VSPGYGVTPNLNVPFSTNVKYYD"
     CDS              2839. .4479
                      /note="GeneID exon scores (in order of location ranges):
                      88.56 - GSCJ_ID dd_01260"
                      /codon_start=1
                      /product="hypothetical protein"
                      /protein_id="AAO51641.2"
                      /db_xref="GI:42733692"
                      /translation="MDCKIHNEKNIMWCLNCTTLCCVDCIEFDHFQHNLKSFKNIDKF
                      HCPTHDEPIIKYCILCNCVICNLCEKTTHFHNNNNNQINITQFPWIPKKKSKSIELEL
                      NEIDEYNNNYNNNINNNNNNNNKNNVNNVNNVNNENKEINENHNNKNQRSNSPPPLPK
                      IKLLNIENITLIKHGSYRNLNKIEEENENDEKEILDRKQYNKENKEFLKKQSEKGNSI
                      AQCLLGSSYYYGMNGYSVNFEKSLKYLTKSAEQNNAKAYFELFNYYKQQKKDLIKSKY
                      YLLKSAEFIYEDSIILLGDYYYNGFKTIDNHDDHDDGSGENEELNLICQDYFKAFNLF
                      KIGASLNVDSNSLTKLGICFYYGRGVTIDYNEAYRLFYQSSICDHHQEHNPEITFYYL
                      GLCLFYGKGVLKNQCKGFEYFMKSASLNHYSPALEAVGRCFLNGEGISQNFMQAKLYF
                      TTAKSQGSNQDRNIQDVNKIIELNKSIKIKSPNILIEIEKIINENKSFNLEQSCLNYI
                      NHQKNNYKQEIDNTNIKNIKCKQDSLIDKTFNKVNYLN"
     CDS              complement(join(4673. .5239,5354. .6118,6205. .6621,
                      6720. .7982))
                      /note="GeneID exon scores (in order of location ranges):
                      34.62, 63.42, 31.78, 100.55 - GSCJ_ID dd_01261"
                      /codon_start=1
                      /product="hypothetical protein"
                      /protein_id="AAO51642.2"
                      /db_xref="GI:42733693"
                      /translation="MEIDKNVEQSSNKVEYNQPKSPITSILVNNNSSNNNNNSNNKNN
                      NNEEEFDISSPLTLSPIANCGASGSNIDSSIDNTPHLLSQISTPPISSTSSITLLNDK
                      KKGIYSTFLDLDYYNNTYLKNKQKILQPKLNQQKQQQQQQQQQQPKTLTFSLDSQQQQQ
                      QQPQLQSPQLIDQQQQQQQQQQQQQQQQQQQQQQQQQIPYHQRKHHISIRDGDDNDNEVDS
                      FNILSQNQLIHQTFNNPQNHHYFSSTSPSLLYHHQQQTQLEQKPIYSSSSSSSILSSL
                      SSSTISFYNSITSLFLTFPQTISTDPIIIREYERLKMLQEQQVGPRHTEILLRKEDGV
                      GGRKALKEKFVEIYESFFNGEDPSENQPIFWEQLFLIKVNIPFLERCIILTSEENLLA
                      LKQNINKIFTQSCFWLKDSNSLRLTHIIETLSILLKSIFRKKFNNFGFDILNILCGIE
                      NSDQVFTDMIKDLEQLLKNSSIDKKVKIGVINLFTIIATATEYINQNTLLQYFMSVNI
                      FDTLVTFLGEDDIPKQTQLNILTLLCYFSNFQKYEITNPYLKDLSNLSDPIKILKILN
                      LITQNLKDHNEFFEQQYRDPNLGMVSRFSGYLSSWVYSPKVVQPLSFYETGSSLMILY
                      ELFYQNDHFVNMLVKGGLPFANGGENGSTCPDIIKQFFSFCAYLASDAQKDLKTIYSR
                      MALLILLCICEKRELEDFFHDIKSTCYIFIYSKKSLFSDPKETEKHPLSCWTLDILSQ
                      FIKCNLKGKISMDVHQSAVDCIHRIVSYKKKTLNRLPYKWSELWSALFSLISMVSTLN
                      ISNDKIQLNQSIHVGISTVNIFNLFINYGDSFLPSPNDYDELFYEIIRSGQVIENFYQ
                      FILNQTTTGGQSVTSPSAAASAFNPTTPTPLSPQQLPTITTTTTTTITSNNESNNPLLN
                      QLLNIRSIVQHFTSKLEEWSANNAEVALTAPQVSKIIKDNYDTLRLKLQENLDQFEPY
                      VENPKEVLFFKHIIKELINDLRKQINVTNII"
     CDS              complement(9007. .10431)
                      /note="GeneID exon scores (in order of location ranges):
                      117.93 - GSCJ_ID dd_03414"
                      /codon_start=1
                      /product="similar to Plasmodium falciparum. Hypothetical
                      protein"
                      /protein_id="AAS38634.1"
                      /db_xref="GI:42733694"
                      /translation="MESYSPSIQKLCVFLKPESDISNTGISYLNNFLADMSNAIIVES
                      IHSSKSHKHKTILIEDIESALNVVFPSEIADSIINQSKGRIQSTSLKNLMQVKTFNNG
                      GNNNNNNEQDNDLQEKEQNELSDDSDVEGFEIIPNNNNIYSHDVSLCFPIDLVCRVTL
                      DYICRKKVIHQNPSDFYKVIEQQQSNHVIYTYISYALEVITTKILIKISSSTDLIDED
                      IIRKVIQDNHHLNSVRHKNIQASNIAGKLLDSSSSSNSNLQTSLSSSFIFSSSSNNTT
                      TFTQQEQQEQFNDDKISGDEIEFEINNSKYLEKEDKKDQLIQKDEEIILKTFNDNNNN
                      NNNNNNNNNNNNNNNNNININNDNDNNNNNNKNNDNKNYDNNNNNNADKIEKLSPVN
                      NNVKNSEIEKNIEKLKNINNNNNVEINEALISIIKETCKNIPSSSFSSVKNLKRSPPS
                      LFKFKEPIDLVQVLAEGGIHMEAT"
     CDS              complement(11849. .12583)
                      /note="GeneID exon scores (in order of location ranges):
                      50.22 - GSCJ_ID dd_02577"
                      /codon_start=1
```

```
                     /product="hypothetical protein"
                     /protein_id="AAS38635.1"
                     /db_xref="GI:42733695"
                     /translation="MKLLFGLILLIVNYCYLINANKLFVNFVGTYNSESCSESPIGTG
                     FSMALGQCVQMQGIFSYVTSSSPEATIWMATLGGDGNSFTLKQFHSSDKLCTEDPINT
                     VKYNKFDTCVQQPSFLTSNNSLLANNNNNNNNNNNNNNNNNNNNLVYSKLTLSQNAPS
                     YAPNSVIIARYNSSDSVCNYNNQNYGISYSTGSVIYNENGVSVYVVCDQSNQSILFHC
                     QNDQCVQSFTDTSCDMPPNDLQIYCN"
     CDS             complement(13151. .13840)
                     /note="GeneID exon scores (in order of location ranges):
                     57.23 - GSCJ_ID dd_02576"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAS38636.1"
                     /db_xref="GI:42733696"
                     /translation="MKLLLTLILIIVNYCCFINANKLFVNFVGSYSSNTCDGNPSGIG
                     FSLAVGQCLPIEGVLSYVINDMPPTTNWFITMGGNGNSFTLKQYDSQDQWCKQDPINT
                     LQFDNFDTCVEQPAFINVNNSLLADTPVVYSKLTLSQNVPTYAPDSVIIGQYNTNTNV
                     CTYNDQQFGVSYTSGTVIYNENGISIHVLCNQVNQSMMFKCKESKCGETFIDTYCQMS
                     GDVQIYCNSNN"
     CDS             join(20802. .21095,21199. .21978)
                     /note="GeneID exon scores (in order of location ranges):
                     23.32, 97.21 - GSCJ_ID dd_01856"
                     /codon_start=1
                     /product="similar to Dictyostelium discoideum (Slime
                     mold). Guanine nucleotide-binding protein alpha-2 subunit
                     (G alpha 2)"
                     /protein_id="AAS38637.1"
                     /db_xref="GI:42733697"
                     /translation="MGICASSMEGEKTNTDINLSIEKERKKKHNEVKLLLLGAGESGK
                     STISKQMKIIHQSGYSNEERKEFKPIITRNILDNMRVLLDGMGRLGMTIDPSNSDAAV
                     MIKELTSLQASIVTDCWGELNEDQGKKIKALWTDPGVKQAMRRANEFSTLPDSAPYFF
                     DSIDRMTSPVYIPTDQDILHTRVMTRGVHETNFEIGKIKFRLVDVGGQRSERKKWLSC
                     FDDVTAVVFCVALSEYDLLLYEDNSTNRMLESLRVFSDVCNSWFVNTPIILFLNKSDL
                     FREKIKHVDLSETFPEYKGGRDYERASNYIKERFWQINKTEQKAIYSHITCATDTNNI
                     RVVFEAVKDIIFTQCVMKAGLYS"
     CDS             join(22686. .22743,22820. .22926,23004. .23139,23232. .23558,
                     23639. .27180)
                     /note="GeneID exon scores (in order of location ranges):
                     -4.93, 3.57, -1.17, 18.69, 174.64 - GSCJ_ID dd_01857"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAS38638.1"
                     /db_xref="GI:42733698"
                     /translation="MNRNIHLYIFVSILLSLINGIKSIKVGNGDFSVIDITDSFDDLV
                     TLYNDQYSIGGLDENIFWLKVINTVQIGENQKFNLLLTLDNSTTIQFDDFFTLTCKER
                     PVNIERELVGQKLFIDKQYKGRYGILVHNIGLEQYLSVPGLENNFISNTFTQGYYDDS
                     IFYTDVSKNYTIFKPVIKRSASVISQLSIDTIFGSVSTSTKLYLTSFLNSTYFKNPTI
                     APLQHDNNCSSLVYDQCSLLFVVQADFQRPGIYFLMNMDSNSNLPELVYGNKSYGHWL
                     FTIKNIAERQGSYIYPISSFDSTLGFSTIFSRNISYFGPTGVNLPVIQQSYFGTQIQN
                     SIFISFSKFVIPTTVNYGDIFIDNGFNTLSHFKYPYFINTINAQGGATNTLWVPSSPY
                     QKSIALEITLQFYFIKYPSPILVGGISGIASFKESIPPVFENIEYFDSVSDDCIVRIS
                     ASDEGGSCIQSIAISFIGAYEPSKVQSIFRITNRNLIEGDCFKGIYQISIPKYSIRRN
                     YFIEMNDYANNLVKLYPNNFYSGKRYLGSFPYQVLTLEDIESIGFRNNNIDLSNSNSS
                     NVLYIKLKDQSISNLVFQFIPKLGIDTYEPNNKISGTPRFFKSKWNSASSIYEVSFDL
                     PARLPTGPISYILNPLEIDSSILYQYFGSPSELNVYSQIADIMGPIALSYNTIGNTTI
                     PSINEPPNTIGFQILFGDYLNGISQVNITVTSDKTYKQMFKNFTFTEPSLGKWIQFTW
                     EIKYGDIPQTFSISELSAIDSNGYITEFPNPTKVNPLMLLVELYPSFSIQTSGETPTD
                     VTPPYMQVYHVNQLNIENYLDRTLEINFTVYDGESGIYIEEPPTIYFINDNVEIISKQ
                     TSYISGTSSGLGFFENAIFGVNITLPYLFGYPNGFLTQIHGIYDNKFNVLGATPDIIN
                     QYNRGNSFINVTIGQSQPPIESFVMDIDGDLLTVNGLKFGILPTGKLRINGAIENLNI
                     SLPFLNNTYIVFDLKQYRNINDTNCQIYITNQLGFSSNEINFVLKAHTNVTMPPTTTP
                     HPTSNPTATPSLTANPTPLPTNSPSQCKNDCGGIAQGQCSNLGCICISPYIGIDCSSK
                     VIVDIDPIVNNSQPEVITPYQHENSGNKFSTLVSIYSLRELGFDGKVENDYIFEMWNY
                     QYYNNDNHTYFTSIKNKNDLSKRTNITVSIKMYKQLTTIQFAGENITIQPSSIKYFIS
                     LSDYQFKSNLNTLDLVMYTSIETEKIDDACFSSEFGDSSSQLSSQYLKIQINDVSLYG
                     EFIKRAIVDTHIVGVLNKVRGNNINSTSNGSSSSQYISIQIPNYKTSVELDPNFSVLL
                     ESSKIENSTCNKNQSGLTKTQLIGIIIGSTAGFILILIIVISIVYKKSTTLRININRA
                     KKFRMKKLSFK"
     CDS             complement(join(27340. .29056,29330. .29490))
                     /note="GeneID exon scores (in order of location ranges):

Alignment Scores:

Pred. No.:              1.75e+04        Length:         88549
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-858-332G-28 (1-6) x AC116924 (1-88549)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      46080 CACAATCACAATCACAAT 46063

RESULT 524
AP007347
LOCUS       AP007347               88709 bp    DNA     linear   HTG 28-DEC-2004
DEFINITION  Lotus corniculatus var. japonicus chromosome 4 clone LjT09M24, ***
            SEQUENCING IN PROGRESS ***, 12 unordered pieces.
ACCESSION   AP007347
VERSION     AP007347.1  GI:56805663
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1
  AUTHORS   Kaneko,T., Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Structural Analysis of a Lotus japonicus Genome. XI. Sequence
            Features and Mapping of Nine hundred twenty-one TAC Clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 88709)
  AUTHORS   Sato,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-2004) Shusei Sato, Kazusa DNA Research Institute,
            Department of Plant Gene Research; 2-6-7 Kazusa-kamatari, Kisarazu,
            Chiba, 292-0818, Japan (E-mail:ssato@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/, Tel:81-438-52-3935(ex.2337),
            Fax:81-438-52-3934)
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     7363: contig of 7363 bp in length
            *     7364     7463: gap of unknown length
            *     7464     8431: contig of 968 bp in length
            *     8432     8531: gap of unknown length
            *     8532    10396: contig of 1865 bp in length
            *    10397    10496: gap of unknown length
            *    10497    11898: contig of 1402 bp in length
            *    11899    11998: gap of unknown length
            *    11999    15928: contig of 3930 bp in length
            *    15929    16028: gap of unknown length
            *    16029    22301: contig of 6273 bp in length
            *    22302    22401: gap of unknown length
            *    22402    33638: contig of 11237 bp in length
            *    33639    33738: gap of unknown length
            *    33739    42651: contig of 8913 bp in length
            *    42652    42751: gap of unknown length
            *    42752    53224: contig of 10473 bp in length
            *    53225    53324: gap of unknown length
            *    53325    67424: contig of 14100 bp in length
            *    67425    67524: gap of unknown length
            *    67525    80914: contig of 13390 bp in length
            *    80915    81014: gap of unknown length
            *    81015    88709: contig of 7695 bp in length.
FEATURES             Location/Qualifiers
     source          1. .88709
```

```
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="genomic DNA"
                     /variety="japonicus"
                     /db_xref="taxon:34305"
                     /chromosome="4"
                     /clone="LjT09M24"
                     /clone_lib="LjT library"
                     /note="TAC clone:TM0432, synonym:Lotus japonicus"
     gap             7364. .7463
                     /estimated_length=unknown
     gap             8432. .8531
                     /estimated_length=unknown
     gap             10397. .10496
                     /estimated_length=unknown
     gap             11899. .11998
                     /estimated_length=unknown
     gap             15929. .16028
                     /estimated_length=unknown
     gap             22302. .22401
                     /estimated_length=unknown
     gap             33639. .33738
                     /estimated_length=unknown
     gap             42652. .42751
                     /estimated_length=unknown
     gap             53225. .53324
                     /estimated_length=unknown
     gap             67425. .67524
                     /estimated_length=unknown
     gap             80915. .81014
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:                1.76e+04        Length:         88709
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       14              Gaps:           0

US-09-858-332G-28 (1-6) x AP007347 (1-88709)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       9889 CATAATCATAATCATAAC 9906

RESULT 525
AP007564
LOCUS       AP007564               88929 bp    DNA     linear   HTG 28-DEC-2004
DEFINITION  Lotus corniculatus var. japonicus clone LjT03D12, *** SEQUENCING IN
            PROGRESS ***, 16 unordered pieces.
ACCESSION   AP007564
VERSION     AP007564.1  GI:56805878
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1
  AUTHORS   Kaneko,T., Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Structural Analysis of a Lotus japonicus Genome. XI. Sequence
            Features and Mapping of Nine hundred twenty-one TAC Clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 88929)
  AUTHORS   Sato,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-2004) Shusei Sato, Kazusa DNA Research Institute,
            Department of Plant Gene Research; 2-6-7 Kazusa-kamatari, Kisarazu,
            Chiba, 292-0818, Japan (E-mail:ssato@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/, Tel:81-438-52-3935(ex.2337),
            Fax:81-438-52-3934)
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 16 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     6673: contig of 6673 bp in length
            *     6674     6773: gap of unknown length
            *     6774     8187: contig of 1414 bp in length
            *     8188     8287: gap of unknown length
            *     8288     9207: contig of 920 bp in length
            *     9208     9307: gap of unknown length
            *     9308    11320: contig of 2013 bp in length
            *    11321    11420: gap of unknown length
            *    11421    13154: contig of 1734 bp in length
            *    13155    13254: gap of unknown length
            *    13255    15374: contig of 2120 bp in length
            *    15375    15474: gap of unknown length
            *    15475    17175: contig of 1701 bp in length
            *    17176    17275: gap of unknown length
            *    17276    20934: contig of 3659 bp in length
            *    20935    21034: gap of unknown length
            *    21035    25527: contig of 4493 bp in length
            *    25528    25627: gap of unknown length
            *    25628    30412: contig of 4785 bp in length
            *    30413    30512: gap of unknown length
            *    30513    38072: contig of 7560 bp in length
            *    38073    38172: gap of unknown length
            *    38173    44759: contig of 6587 bp in length
            *    44760    44859: gap of unknown length
            *    44860    54467: contig of 9608 bp in length
            *    54468    54567: gap of unknown length
            *    54568    65260: contig of 10693 bp in length
            *    65261    65360: gap of unknown length
            *    65361    85247: contig of 19887 bp in length
            *    85248    85347: gap of unknown length
            *    85348    88929: contig of 3582 bp in length.
FEATURES             Location/Qualifiers
     source          1. .88929
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="genomic DNA"
                     /variety="japonicus"
                     /db_xref="taxon:34305"
                     /clone="LjT03D12"
                     /clone_lib="LjT library"
                     /note="TAC clone:TM0746, synonym:Lotus japonicus"
     gap             6674. .6773
                     /estimated_length=unknown
     gap             8188. .8287
                     /estimated_length=unknown
     gap             9208. .9307
                     /estimated_length=unknown
     gap             11321. .11420
                     /estimated_length=unknown
     gap             13155. .13254
                     /estimated_length=unknown
     gap             15375. .15474
                     /estimated_length=unknown
     gap             17176. .17275
                     /estimated_length=unknown
     gap             20935. .21034
                     /estimated_length=unknown
     gap             25528. .25627
                     /estimated_length=unknown
     gap             30413. .30512
                     /estimated_length=unknown
     gap             38073. .38172
                     /estimated_length=unknown
     gap             44760. .44859
                     /estimated_length=unknown
     gap             54468. .54567
```

```
                     /estimated_length=unknown
     gap             65261. .65360
                     /estimated_length=unknown
     gap             85248. .85347
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              1.76e+04        Length:         88929
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AP007564 (1-88929)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      58919 CATAATCACAACCATAAT 58936

RESULT 526
AP007683
LOCUS       AP007683               89275 bp    DNA     linear   HTG 28-DEC-2004
DEFINITION  Lotus corniculatus var. japonicus chromosome 2 clone LjB19D18, ***
            SEQUENCING IN PROGRESS ***, 7 unordered pieces.
ACCESSION   AP007683
VERSION     AP007683.1  GI:56805993
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1
  AUTHORS   Kaneko,T., Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Structural Analysis of a Lotus japonicus Genome. XI. Sequence
            Features and Mapping of Nine hundred twenty-one TAC Clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 89275)
  AUTHORS   Sato,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-2004) Shusei Sato, Kazusa DNA Research Institute,
            Department of Plant Gene Research; 2-6-7 Kazusa-kamatari, Kisarazu,
            Chiba, 292-0818, Japan (E-mail:ssato@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/, Tel:81-438-52-3935(ex.2337),
            Fax:81-438-52-3934)
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3369: contig of 3369 bp in length
            *     3370     3469: gap of unknown length
            *     3470     4416: contig of 947 bp in length
            *     4417     4516: gap of unknown length
            *     4517    15956: contig of 11440 bp in length
            *    15957    16056: gap of unknown length
            *    16057    30649: contig of 14593 bp in length
            *    30650    30749: gap of unknown length
            *    30750    50820: contig of 20071 bp in length
            *    50821    50920: gap of unknown length
            *    50921    83049: contig of 32129 bp in length
            *    83050    83149: gap of unknown length
            *    83150    89275: contig of 6126 bp in length.
FEATURES            Location/Qualifiers
     source         1. .89275
                    /organism="Lotus corniculatus var. japonicus"
                    /mol_type="genomic DNA"
                    /variety="japonicus"
                    /db_xref="taxon:34305"
                    /chromosome="2"
                    /clone="LjB19D18"
                    /clone_lib="LjB library"
                    /note="BAC clone:BM0925, synonym:Lotus japonicus"
     gap            3370. .3469
                    /estimated_length=unknown
     gap            4417. .4516
                    /estimated_length=unknown
     gap            15957. .16056
                    /estimated_length=unknown
     gap            30650. .30749
                    /estimated_length=unknown
     gap            50821. .50920
                    /estimated_length=unknown
     gap            83050. .83149
                    /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              1.77e+04        Length:         89275
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AP007683 (1-89275)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       9500 CATAACCATAACCATAAC 9517

RESULT 527
AC015216/c
LOCUS       AC015216               89387 bp    DNA     linear   HTG 16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC015216
VERSION     AC015216.1  GI:6436119
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 89387)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10212548 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES            Location/Qualifiers
     source         1. .89387
                    /organism="Drosophila melanogaster"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:7227"
ORIGIN

Alignment Scores:
Pred. No.:              1.77e+04        Length:         89387
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AC015216 (1-89387)
```

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      57686 CACAATCACAATCACAAT 57669

RESULT 528
AC166640/c
LOCUS       AC166640               89426 bp    DNA     linear   HTG 03-AUG-2005
DEFINITION  Solanum demissum chromosome 11 clone PGEC48G14, *** SEQUENCING IN
            PROGRESS ***.
ACCESSION   AC166640
VERSION     AC166640.1  GI:71725489
KEYWORDS    HTG; HTGS_PHASE2; HTGS_ACTIVEFIN.
SOURCE      Solanum demissum
  ORGANISM  Solanum demissum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum.
REFERENCE   1  (bases 1 to 89426)
  AUTHORS   Buell,R., Lee,D., Tallon,L.J., Kuang,H. and Baker,B.
  TITLE     Solanum demissum BAC genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 89426)
  AUTHORS   Buell,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-2005) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1    89426: contig of 89426 bp in length.
FEATURES            Location/Qualifiers
     source         1. .89426
                    /organism="Solanum demissum"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:50514"
                    /chromosome="11"
                    /clone="PGEC48G14"
ORIGIN

Alignment Scores:
Pred. No.:              1.77e+04        Length:         89426
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AC166640 (1-89426)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      25003 CACAATCACAATCACAAT 24986

RESULT 529
AP007566/c
LOCUS       AP007566               89853 bp    DNA     linear   HTG 28-DEC-2004
DEFINITION  Lotus corniculatus var. japonicus chromosome 2 clone LjT13J06, ***
            SEQUENCING IN PROGRESS ***, 23 unordered pieces.
ACCESSION   AP007566
VERSION     AP007566.1  GI:56805880
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1
  AUTHORS   Kaneko,T., Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Structural Analysis of a Lotus japonicus Genome. XI. Sequence
            Features and Mapping of Nine hundred twenty-one TAC Clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 89853)
  AUTHORS   Sato,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-2004) Shusei Sato, Kazusa DNA Research Institute,
            Department of Plant Gene Research; 2-6-7 Kazusa-kamatari, Kisarazu,
            Chiba, 292-0818, Japan (E-mail:ssato@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/, Tel:81-438-52-3935(ex.2337),
            Fax:81-438-52-3934)
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 23 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     4255: contig of 4255 bp in length
            *    4256     4355: gap of unknown length
            *    4356     5370: contig of 1015 bp in length
            *    5371     5470: gap of unknown length
            *    5471     6262: contig of 792 bp in length
            *    6263     6362: gap of unknown length
            *    6363     7363: contig of 1001 bp in length
            *    7364     7463: gap of unknown length
            *    7464     8654: contig of 1191 bp in length
            *    8655     8754: gap of unknown length
            *    8755     9685: contig of 931 bp in length
            *    9686     9785: gap of unknown length
            *    9786    11120: contig of 1335 bp in length
            *   11121    11220: gap of unknown length
            *   11221    12896: contig of 1676 bp in length
            *   12897    12996: gap of unknown length
            *   12997    15021: contig of 2025 bp in length
            *   15022    15121: gap of unknown length
            *   15122    17056: contig of 1935 bp in length
            *   17057    17156: gap of unknown length
            *   17157    19168: contig of 2012 bp in length
            *   19169    19268: gap of unknown length
            *   19269    21456: contig of 2188 bp in length
            *   21457    21556: gap of unknown length
            *   21557    24275: contig of 2719 bp in length
            *   24276    24375: gap of unknown length
            *   24376    26905: contig of 2530 bp in length
            *   26906    27005: gap of unknown length
            *   27006    32527: contig of 5522 bp in length
            *   32528    32627: gap of unknown length
            *   32628    38257: contig of 5630 bp in length
            *   38258    38357: gap of unknown length
            *   38358    43750: contig of 5393 bp in length
            *   43751    43850: gap of unknown length
            *   43851    48308: contig of 4458 bp in length
            *   48309    48408: gap of unknown length
            *   48409    53465: contig of 5057 bp in length
            *   53466    53565: gap of unknown length
            *   53566    58723: contig of 5158 bp in length
            *   58724    58823: gap of unknown length
            *   58824    70624: contig of 11801 bp in length
            *   70625    70724: gap of unknown length
            *   70725    81783: contig of 11059 bp in length
            *   81784    81883: gap of unknown length
            *   81884    89853: contig of 7970 bp in length.
FEATURES            Location/Qualifiers
     source         1. .89853
                    /organism="Lotus corniculatus var. japonicus"
                    /mol_type="genomic DNA"
                    /variety="japonicus"
                    /db_xref="taxon:34305"
```

```
                     /chromosome="2"
                     /clone="LjT13J06"
                     /clone_lib="LjT library"
                     /note="TAC clone:TM0749, synonym:Lotus japonicus"
     gap             4256. .4355
                     /estimated_length=unknown
     gap             5371. .5470
                     /estimated_length=unknown
     gap             6263. .6362
                     /estimated_length=unknown
     gap             7364. .7463
                     /estimated_length=unknown
     gap             8655. .8754
                     /estimated_length=unknown
     gap             9686. .9785
                     /estimated_length=unknown
     gap             11121. .11220
                     /estimated_length=unknown
     gap             12897. .12996
                     /estimated_length=unknown
     gap             15022. .15121
                     /estimated_length=unknown
     gap             17057. .17156
                     /estimated_length=unknown
     gap             19169. .19268
                     /estimated_length=unknown
     gap             21457. .21556
                     /estimated_length=unknown
     gap             24276. .24375
                     /estimated_length=unknown
     gap             26906. .27005
                     /estimated_length=unknown
     gap             32528. .32627
                     /estimated_length=unknown
     gap             38258. .38357
                     /estimated_length=unknown
     gap             43751. .43850
                     /estimated_length=unknown
     gap             48309. .48408
                     /estimated_length=unknown
     gap             53466. .53565
                     /estimated_length=unknown
     gap             58724. .58823
                     /estimated_length=unknown
     gap             70625. .70724
                     /estimated_length=unknown
     gap             81784. .81883
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              1.78e+04        Length:         89853
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AP007566 (1-89853)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      44860 CATAACCACAATCATAAC 44843

RESULT 530
AP007932
LOCUS       AP007932               90264 bp    DNA     linear   HTG 28-DEC-2004
DEFINITION  Lotus corniculatus var. japonicus chromosome 1 clone LjT16L24, ***
            SEQUENCING IN PROGRESS ***, 15 unordered pieces.
ACCESSION   AP007932
VERSION     AP007932.1  GI:56806240
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1
  AUTHORS   Kaneko,T., Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Structural Analysis of a Lotus japonicus Genome. XI. Sequence
            Features and Mapping of Nine hundred twenty-one TAC Clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 90264)
  AUTHORS   Sato,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-2004) Shusei Sato, Kazusa DNA Research Institute,
            Department of Plant Gene Research; 2-6-7 Kazusa-kamatari, Kisarazu,
            Chiba, 292-0818, Japan (E-mail:ssato@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/, Tel:81-438-52-3935(ex.2337),
            Fax:81-438-52-3934)
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 15 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     4404: contig of 4404 bp in length
            *     4405     4504: gap of unknown length
            *     4505     5535: contig of 1031 bp in length
            *     5536     5635: gap of unknown length
            *     5636     6672: contig of 1037 bp in length
            *     6673     6772: gap of unknown length
            *     6773     8063: contig of 1291 bp in length
            *     8064     8163: gap of unknown length
            *     8164     9254: contig of 1091 bp in length
            *     9255     9354: gap of unknown length
            *     9355    10365: contig of 1011 bp in length
            *    10366    10465: gap of unknown length
            *    10466    13173: contig of 2708 bp in length
            *    13174    13273: gap of unknown length
            *    13274    16189: contig of 2916 bp in length
            *    16190    16289: gap of unknown length
            *    16290    22227: contig of 5938 bp in length
            *    22228    22327: gap of unknown length
            *    22328    30573: contig of 8246 bp in length
            *    30574    30673: gap of unknown length
            *    30674    41038: contig of 10365 bp in length
            *    41039    41138: gap of unknown length
            *    41139    55026: contig of 13888 bp in length
            *    55027    55126: gap of unknown length
            *    55127    70146: contig of 15020 bp in length
            *    70147    70246: gap of unknown length
            *    70247    87632: contig of 17386 bp in length
            *    87633    87732: gap of unknown length
            *    87733    90264: contig of 2532 bp in length.
FEATURES             Location/Qualifiers
     source          1. .90264
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="genomic DNA"
                     /variety="japonicus"
                     /db_xref="taxon:34305"
                     /chromosome="1"
                     /clone="LjT16L24"
                     /clone_lib="LjT library"
                     /note="TAC clone:TM1253, synonym:Lotus japonicus"
     gap             4405. .4504
                     /estimated_length=unknown
     gap             5536. .5635
                     /estimated_length=unknown
     gap             6673. .6772
                     /estimated_length=unknown
     gap             8064. .8163
                     /estimated_length=unknown
```

```
     gap             9255. .9354
                     /estimated_length=unknown
     gap             10366. .10465
                     /estimated_length=unknown
     gap             13174. .13273
                     /estimated_length=unknown
     gap             16190. .16289
                     /estimated_length=unknown
     gap             22228. .22327
                     /estimated_length=unknown
     gap             30574. .30673
                     /estimated_length=unknown
     gap             41039. .41138
                     /estimated_length=unknown
     gap             55027. .55126
                     /estimated_length=unknown
     gap             70147. .70246
                     /estimated_length=unknown
     gap             87633. .87732
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              1.79e+04        Length:         90264
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AP007932 (1-90264)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      79737 CATAACCACAACCACAAT 79754

RESULT 531
AC127632_3/c
WPCOMMENT
  Sequence split into 4 fragments  LOCUS AC127632 Accession AC127632
     Fragment Name      Begin     End
     AC127632_0             1     110000
     AC127632_1        100001     210000
     AC127632_2        200001     310000
     AC127632_3        300001     390275
Continuation (4 of 4) of AC127632 from base 300001 (AC127632 Rattus norvegicus clone CH2

Alignment Scores:
Pred. No.:              1.79e+04        Length:         90275
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AC127632_3 (1-90275)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       2950 CACAATCACAACCACAAC 2933

RESULT 532
AC115680
LOCUS       AC115680               90373 bp    DNA     linear   INV 21-FEB-2004
DEFINITION  Dictyostelium discoideum chromosome 2 map 4915084-5005461 strain
            AX4, complete sequence.
ACCESSION   AC115680
VERSION     AC115680.3  GI:42733611
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 90373)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 90373)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 90373)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 90373)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   5  (bases 1 to 90373)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2004) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On Feb 21, 2004 this sequence version replaced gi:28828453.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES            Location/Qualifiers
     source         1. .90373
                    /organism="Dictyostelium discoideum"
                    /mol_type="genomic DNA"
                    /strain="AX4"
                    /db_xref="taxon:44689"
                    /chromosome="2"
                    /map="4915084-5005461"
     CDS            2635. .4032
                    /note="GeneID exon scores (in order of location ranges):
                    101.29 - GSCJ_ID dd_02405"
                    /codon_start=1
                    /product="similar to Plasmodium falciparum. Hypothetical
                    protein"
                    /protein_id="AAO51072.2"
                    /db_xref="GI:42733612"
                    /translation="MVNEHNVTLSNLIDFGLIKPNQEVKYSYRGVSYTGVILLNGEIH
                    TNGVSFTNPTHWTRTISGNNCSGWGTVKLSGASGPPLLKLKREYLFRVGSTGKNKKKT
                    QKQQQPQQPQPQPQPQQQQQQQQQQQQQQQQQQQQPNNNDNDSESEIYKKSDTSETSDQ
                    DIDNDDDDAGGQSNNRTTTTTTTTTTTTTTNRQNSPQKKIPSTAITNPKEKKKEKKEN
                    ILTKKKQQSLQYQQQLQLLQRQNSPPSVSPSSSTSTSSSTSSPASNQIFNSFGPNSQN
                    HNQYGINYNSQQHQPQQYNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
                    NNNNNNNSNNNNNNNNNNNNININNNNNINNNNNINNNSNNQPNCCSPNSNCQYYGY
                    NQQNPPKHQYPNNFLTDGHNPSPPIPVLPFTNQSQNQNQNQNQNQNQKSKSKSKSK
                    SKFKSKSKSKSKSKS"
     CDS            4130. .6088
                    /note="GeneID exon scores (in order of location ranges):
                    190.41 - GSCJ_ID dd_02404"
                    /codon_start=1
                    /product="similar to Dictyostelium discoideum (Slime
                    mold). Prespore-specific protein (Fragment)"
                    /protein_id="AAS38587.1"
                    /db_xref="GI:42733613"
                    /translation="MYQQQQQQQQQQQQHNQQQLLSHQHPQQPNYPPQYNYHGQINHS
                    QQQQQQQQQQYQESTNPPFTPVSMPYGVFQSPLYGLPNIEQDDMYGGYPPKYSTSQNL
                    LPGTPQANNNFINCFNQTFPSRGYQSQPHNLFLPQQPQQQPSPLMSQQLPFGNQQASI
```

```
                SSPTSTSSPQLSQFYSNYQQFQLQMHHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ
                QQQQQQQQQQTIQQQQTIQQQQTIQQQLQQHQQHQLQHQQHQQQHQQQQHQQQQHQLQ
                QQQQHQQQQQQQQLQQQQQQQQQLQQQQLQHETIQQQIQQQHKQIQQQQQQQNQSLSSP
                TSSDAANSSPSIFDKKRKGKEIEKEKKPTVKKETDSKKQSSQTKKPSPTSTFKKAKLS
                NIKIDNNNNNPSNNNFNNSNINNSTNNNSNINNSTNNNSTNNNNNNNNNNNNNNN
                NNFNNNNSNNNNNSNINNSNNNNSNNTYNGNNSFGNNNSANSNQYSPSSYNAQLPNNN
                NNNNNNSNNTNNNSNNNNISQNYVSNSSQEPSSNCNIISTPINNQEVIGYNSGAFNNN
                NNNNNGDGITHHHQQYQLHPQQLQQQQQQQQQPQEQQYRQIPQQYEQYQEYQEPTQYQ
                QLYQYQYHSEQQQQHQTEVQQFNQDTNQ"
CDS             complement(join(6307. .6461,6544. .6634))
                /note="GeneID exon scores (in order of location ranges):
                10.25, 9.85 - GSCJ_ID dd_02403"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAS38588.1"
                /db_xref="GI:42733614"
                /translation="METFEKRANEADEIIEQLTKRISQLEQSILNKKKQTIVPTSSLS
                ESVPLSTYEALLKENEKLKQDMEKQNYRINHLLKNMK"
CDS             join(7653. .8049,8184. .8444,8516. .8845,8972. .9148,
                9256. .9714,9828. .10175,10265. .12289,12408. .12631)
                /note="GeneID exon scores (in order of location ranges):
                9.95, 10.45, 12.15, 7.00, 15.98, 11.85, 142.36, 13.43 -
                GSCJ_ID dd_01699"
                /codon_start=1
                /product="similar to Dictyostelium discoideum (Slime
                mold). Hypothetical 127.0 kDa protein"
                /protein_id="AAL92634.3"
                /db_xref="GI:42733615"
                /translation="MDDFKNLLIFLIFLFFRLNNGQITPLTPTNSITITPLNDFRFSQ
                FPIISSLGECSTITGKFLVVDDNANNEITKISIIQDKIGSQSLTKEGNHYLFSFWLSL
                NFTVSPLQMNITISSNSTSDVVYSYTQSCVELENPITILAESVPITKFKLKNYSNGST
                LLVFNNGFGSFTALDDGTYLIVWDGKYVSSKTNLYDDWSFDILFYKKNKTFTVPSYFS
                FDKQLYAPTYNKTYYPNDIMDIGSNYTEFGQDYPPVLTNNVKTNIPHPFSGVSDLSYF
                YLQKPLKGKIGNMTYLNLLLCSTKYNSGDPLNFVYYDQYQNQTSKTPFVTKIYKFNVQ
                NLGNIMLSNTTSKSPIIELFTVSFKNVVDYKFTIYQYTYDYKKFFKTWPYGFIKGSNS
                DYTHKTSIQSNMKRLSSYITTNYTIQTFSQSDGATIAVPKSTIPGLKAPTVVSYRFYE
                TYTGFFILSVDFLSENGLENAYSSSLDINPGLLNLVSGSIYNGTLDFIFAINESPKVT
                LNDIYDNTFFLNEFDLLSINPLSQFKYPHQLYYNFNISNFYNISFSANNLNLEKDNYI
                TVYLCNKDLPTDTTFIFYNVDDSSYTAPQLISKTPSTSLIYNSTLGCFTSRLIIQKNS
                IFFEVLGIPYSGQLSFFYSSMLPKSSRLFIKKSELDAQGPIFSFIKKINDKVDVSSNS
                VIVGWNFTISDEGNGFSHGYIKIMGSVDSSTYHYNFTIDDIVEGDKFNGIYQFVISVS
                TPCVSQEYSITNVELYDTNNVISIFELYTSKTFSAYYSNPFINFYQSGIDITKVSFTC
                TPSILDTTDPILKSFTVSKDSIDVGSNDRSIIFNFEIETNSTGTGLKIDQLPIVYILS
                KKMKLLECKSTLLSNNVTNYIYKCQIELPVSFGHPYPISYSLYGIVNNGGNFNGFTTE
                DLKLNGFPYYSNTTFSTDIPVVMSNHVIYSDIGGDLFLRGRGFSNSKNVEITYSQVGL
                TSQNLVPTIISHSIIKVTGIKPTSESILVKVETENYFSNSYTVYPIFFNETIPEPPET
                PTPTQTQTQTQTPTPTQIPTASPLPTNPPQECKGSCGGINKGYCSPTGCICYSPWVGE
                TCSSQVIIVPPPKPDTGFPTTQFPDSINSNSNSSASDSGNTSEENNRPITFNSLISIV
                SLREIDYKGEVVVEHKFEKWYFTPITELKSQYITTINNNNTNVTATLEWFNKSTTIQF
                ANENLEMNPSTIKYSIEISQYNFKQNLNNLQLIMFASFDASDPPKDICSSKQFGNTST
                GDNSNYIKIQIDNHSLYGRFIKRAIIDDKVYSIENQLLDSSLNSISNSSSVQSYIGIT
                IPFYKNYVVVDPDFSVLVDSNSVSSSSTNSMCNSSSSGLSKSQLIGIIIGSSCFAIAI
                VVGLIYLARKRHVQKLFLEKYSSRMKSK"
CDS             complement(join(13117. .13520,13570. .13783,13864. .14029,
                14242. .14702))
                /note="GeneID exon scores (in order of location ranges):
                5.46, 9.65, 4.41, 8.61 - GSCJ_ID dd_01697"
                /codon_start=1
                /product="similar to Plasmodium falciparum (isolate 3D7).
                Hypothetical protein"
                /protein_id="AAL92635.3"
                /db_xref="GI:42733616"
                /translation="MNENVINKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNKLNS
                LVFIDLIEYLLYEQSFERILDVSLLNKFIFKNVQQIITNKQQSIKFFPEMESLLLFIN
                PNHDENIESLQQNDCVGEYRLKLIQYKEIGNLYTQFFGKIKFFHDNNFNVGIPKSIYI
                DSEGDGKEIDPKEVQTILNCKSIKKIKDINNQTPSTFFKESIYSTTTPSLLNNDCESL
                SKTLIHKRRYELIVQLLINSHVSASKLKELVIMASLRNEFIIKIDNTCFLEPLIGNTT
                LTTLEGYTRNPINDITQSSITLSPLINLPTITTLHCELYSLFTFLQHCKENSNIKHLV
                ISSIFYFSKRKCGSDRFIDDEIQFLYDFLKHNKSLQYLEIKLEKVKEKWDLFDNNNNN
                QFDKILSFFKNSNIKTLKQLYY"
CDS             complement(14817. .16577)
                /note="GeneID exon scores (in order of location ranges):
                57.48 - GSCJ_ID dd_03173"
                /codon_start=1
                /product="similar to Plasmodium falciparum (isolate 3D7).
                Hypothetical 111.5 kDa protein"
                /protein_id="AAS38589.1"
                /db_xref="GI:42733617"
                /translation="MNNNKNNNNNNNNNNNNNNNNNNNNNNNNNNKKLKNVNTIKLNS
                LVFIDVLPYLLYNRPFKIILEISLINKSIFKYVQQIITYQQQLIKFFRKINSFIFFIH
                NDQNTKNEIVEPVLKKRNIQTLSFLQNKNHHNSITKNHQHEGEYNYKLIQYKEIENLY
                CTFNREFNSINLNQFDILKNHLIILINDNDNNNNNNNNNNNNNNNNNNNNKKLKNVI
                ITNRVLADTYSKNPTCDGFFKFNKIKLFWDQEIYPCSSTHEYYDYGFRDNYLLFEYSF
                EFLNRYKPKSIYIITTKIGPVPLHIQNLNQILNCETIEEIKFGYHQTPSLFFKNAIKKT
                NETSTNLKSLYLRLPNFKDEIEYDNYDGIDFIKKLLNCNESNDDNNLLMELVKEKKEE
                LVSNLLKECNLSKNKILKELLIFCPIRDDFILDDSSNLLESLIGNETLTTLGLSGFKF
                RKFDEIQTKYIKIPTIEIEESSSSSKTLSPLIKIPTISSLYCEIYSLLPFLQHCNENSN
                IKRLIVVADDPPLFYTNKFSSIIFNGQEIQFISNFLKQNKSLKYLEIKLEIHNVKKKW
                DILNDFEEILGFLSKLRLLE"
CDS             17146. .18276
                /note="GeneID exon scores (in order of location ranges):
                192.42 - GSCJ_ID dd_01695"
                /codon_start=1
                /product="similar to Dictyostelium discoideum (Slime
                mold). Actin 15 (Actin A8)"
                /protein_id="AAS38590.1"
                /db_xref="GI:42733618"
                /translation="MDGEDVQALVIDNGSGMCKAGFAGDDAPRAVFPSIVGRPRHTGV
                MVGMGQKDSYVGDEAQSKRGILTLKYPIEHGIVTNWDDMEKIWHHTFYNELRVAPEEH
                PVLLTEAPLNPKANREKMTQIMFETFNTPAMYVAIQAVLSLYASGRTTGIVMDSGDGV
                SHTVPIYEGYALPHAILRLDLAGRDLTDYMMKILTERGYSFTTTAEREIVRDIKEKLA
                YVALDFEAEMQTAASSSALEKSYELPDGQVITIGNERFRCPEALFQPSFLGMESAGIH
                ETTYNSIMKCDVDIRKDLYGNVVLSGGTTMFPGIADRMNKELTALAPSTMKIKIIAPP
                ERKYSVWIGGSILASLSTFQQMWISKEEYDESGPSIVHRKCF"
CDS             19021. .20070
                /note="GeneID exon scores (in order of location ranges):
                22.99 - GSCJ_ID dd_01704"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAS38591.1"
                /db_xref="GI:42733619"
                /translation="MDSEEHINKKLKIDDIKEENNNNKKLQNYICKNIVELLINYLLK
                SKKKRSNSNNNSFERFEKKQFILYLDEIITLSLVSKKWFKIISDFISNGIIEQKIIDY
                WFSDEGDDATTLHRGEDATTLHRGEDAMKSNKYYSLIKIDATTIDDFFSSKYYYIENF
                QGAKFNPVTTTDDEPKKFINNNNKTFKRIKNDDEPNKFINNNNKTFKRIENDEIKILL
                QITDRLNESVKVDLLYQSIDEEYLIPIEKNDNIIVENLKLEGVDTEYSSTLSNIFNFQ
                PKHIKYNPNGVYCQDSYFHMDYLELFNPDNLEISASPTLPLSTNRTESIKINTTDHVE

Alignment Scores:
Pred. No.:              1.79e+04        Length:         90373
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-858-332G-28 (1-6) x AC115680 (1-90373)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      67687 CACAACCACAACCACAAC 67704

RESULT 533
AC165547/c
LOCUS       AC165547               90572 bp    DNA     linear   HTG 18-JUL-2005
DEFINITION  Bos taurus clone CH240-170G8, *** SEQUENCING IN PROGRESS ***, 25
            unordered pieces.
ACCESSION   AC165547
VERSION     AC165547.2  GI:70912967
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 90572)
```

```
  AUTHORS   Muzny,D., Adams,C., Agbai II,O., Allen,C., Alsbrooks,S., Archer,P.,
            Arredondo,H., Bandaranaike,D., Bangura,L., Beltran,B., Beltran,R.,
            Beraducci,A., Biswalo,K., Blyth,P., Bonham,H., Buhay,C., Burch,P.,
            Cadoree,I., Canada,A., Cardenas,V., Carter,K., Cavazos,I.,
            Chacko,J., Chahrour,M., Chavez,D., Chen,A., Chen,G., Chen,R.,
            Cheng,M.-T., Chu,J., Clerc,K., Cockrell,R., Coyle,M., Cree,A.,
            Curry,S., Dai,W., Davila,M.L., Davis,C., Davy-Carroll,L., De
            Anda,C., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H.,
            Donlin,J., McCauley,S., Dugan-Rocha,S., Dunn,A., Durbin,K.,
            Dziuda,D., Egan,A., Escotto,M., Espinosa,V., Eugene,C., Fa,M.,
            Fernandez,S., Fernando,P., Flagg,N., Forbes,L., Foster,P.,
            Fowler,G., Fu,Q., Fuh,E., Garcia,A., Garcia,R., Garner,T.,
            Gaskin,C., Gench,S., Ghose,S., Gill,R., Gonzalez,D.,
            Gonzalez-Garay,M., Guevara,W., Holder,M., Haaland,W., Haeberlen,K.,
            Hall,B., Hamid,H., Hamilton,K., Harbes,B., Harris,R., Havlak,P.,
            Hawes,A., Hawkins,E., Hayes,S., Hemphill,L., Hernandez,J.,
            Hines,S., Hitchens,M., Hodgson,A., Hogues,M., Hollins,B.,
            Howell,L.T., Hulyk,S., Hume,J., Imo,K., Jackson,A., Jackson,L.,
            Jacob,L., Jiang,H., Johnson,B., Johnson,R., Kalafus,K., Kelly,S.,
            Keys,T., Khan,Z., King,L., Kovar,C., Kowis,A., Kowis,C., Lara,F.,
            Leal,S., Lee,K., Lee,S., LeGall,F.I., Lemon,S., Lewis,L., Li,B.,
            Li,Y., Li,Z., Linnell,M., Liu,W., Liu,Y.-S., Liu,Y., Liyanage,D.,
            London,P., Lopez,J., Lorensuhewa,L., Lozado,R., Luk,T., Madu,R.,
            Maheshwari,M., Mahoney,C., Malloy,K., Mansouri,D., Martinez,E.,
            McClelland,H., McPherson,J., Mercadao,C., Metzker,M.,
            Milosavljevic,A., Minja,E., Morgan,M., Morris,S., Munidasa,M.,
            Murray,D., Nazarith,L., Ngo,D., Nguyen,N., Norwig-Eastaugh,E.,
            Nott,A., Nwaokelemeh,O., Obregon,M., Ochi-Okorie,C., Odeh,E.,
            Okwuonu,G., Okwuonu,K., Parker,D., Pasternak,S., Patel,B.,
            Patel,V., Paul,H., Perez,A., Perez,L., Petrosino,J., Pham,T.,
            Primus,E., Pu,L.-L., Puazo,M., Qin,X., Quinn,A., Quiroz,J.,
            Rabata,D., Rachlin,E., Reigh,R., Ren,Y., Reuter,M., Richards,S.,
            Rives,C., Rodriguez,F., Rojas,A., Ruiz,S.J., Sana,M., Sanders,W.,
            Santibanez,J., Santos,R., Savery,G., Scherer,S., Shen,H., Shen,Y.,
            Sisson,I., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R.,
            Svatek,A., Taylor,E., Taylor,T., Thomas,N., Thorn,R., Thornton,R.,
            Trejos,Z., Usmani,K., Vargo,C., Verduzco,D., Villasana,D., Virk,D.,
            Volkov,A., Waldron,L., Walker,B., Wang,Q., Wang,S., Warren,J.,
            Wei,X., Wheeler,D., Williams,G., Williams,R., Worley,K., Wright,R.,
            Wu,J., Yakub,S., Yan,K., Yuan,Y., Yu,F., Zhang,J., Zhang,L.,
            Zhang,Z., Zhou,J., Weinstock,G. and Gibbs,R.
    TITLE   Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 90572)
  AUTHORS   Worley,K.C.
    TITLE   Direct Submission
  JOURNAL   Submitted (17-JUL-2005) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 90572)
  AUTHORS   Worley,K.C.
    TITLE   Direct Submission
  JOURNAL   Submitted (18-JUL-2005) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 18, 2005 this sequence version replaced gi:70912565.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help.tmc.edu
            -------------- Project Information
                Center project name: FJYT
                Center clone name: CH240-170G8
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 92849 bases at least Q40
                Consensus quality: 96310 bases at least Q30
                Consensus quality: 98411 bases at least Q20
                Estimated insert size: 118858; sum-of-contigs estimation
                Estimated insert size: 22369; agarose-fp estimation
                Quality coverage: 7x in Q20 bases; agarose-fp estimation
                Quality coverage: 1x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 25 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     5510: contig of 5510 bp in length
            *     5511     5610: gap of unknown length
            *     5611     8044: contig of 2434 bp in length
            *     8045     8144: gap of unknown length
            *     8145    10409: contig of 2265 bp in length
            *    10410    10509: gap of unknown length
            *    10510    12634: contig of 2125 bp in length
            *    12635    12734: gap of unknown length
            *    12735    17104: contig of 4370 bp in length
            *    17105    17204: gap of unknown length
            *    17205    20352: contig of 3148 bp in length
            *    20353    20452: gap of unknown length
            *    20453    27166: contig of 6714 bp in length
            *    27167    27266: gap of unknown length
            *    27267    30811: contig of 3545 bp in length
            *    30812    30911: gap of unknown length
            *    30912    33776: contig of 2865 bp in length
            *    33777    33876: gap of unknown length
            *    33877    42113: contig of 8237 bp in length
            *    42114    42213: gap of unknown length
            *    42214    45745: contig of 3532 bp in length
            *    45746    45845: gap of unknown length
            *    45846    49630: contig of 3785 bp in length
            *    49631    49730: gap of unknown length
            *    49731    53881: contig of 4151 bp in length
            *    53882    53981: gap of unknown length
            *    53982    57411: contig of 3430 bp in length
            *    57412    57511: gap of unknown length
            *    57512    62047: contig of 4536 bp in length
            *    62048    62147: gap of unknown length
            *    62148    65501: contig of 3354 bp in length
            *    65502    65601: gap of unknown length
            *    65602    67624: contig of 2023 bp in length
            *    67625    67724: gap of unknown length
            *    67725    69812: contig of 2088 bp in length
            *    69813    69912: gap of unknown length
            *    69913    72864: contig of 2952 bp in length
            *    72865    72964: gap of unknown length
            *    72965    75635: contig of 2671 bp in length
            *    75636    75735: gap of unknown length
            *    75736    78537: contig of 2802 bp in length
            *    78538    78637: gap of unknown length
            *    78638    83388: contig of 4751 bp in length
            *    83389    83488: gap of unknown length
            *    83489    86084: contig of 2596 bp in length
            *    86085    86184: gap of unknown length
            *    86185    88310: contig of 2126 bp in length
            *    88311    88410: gap of unknown length
            *    88411    90572: contig of 2162 bp in length.
FEATURES             Location/Qualifiers
     source          1. .90572
                     /organism="Bos taurus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9913"
                     /clone="CH240-170G8"
     gap             5511. .5610
                     /estimated_length=unknown
     gap             8045. .8144
                     /estimated_length=unknown
     gap             10410. .10509
```

```
                     /estimated_length=unknown
     gap             12635. .12734
                     /estimated_length=unknown
     gap             17105. .17204
                     /estimated_length=unknown
     gap             20353. .20452
                     /estimated_length=unknown
     gap             27167. .27266
                     /estimated_length=unknown
     gap             30812. .30911
                     /estimated_length=unknown
     gap             33777. .33876
                     /estimated_length=unknown
     gap             42114. .42213
                     /estimated_length=unknown
     gap             45746. .45845
                     /estimated_length=unknown
     gap             49631. .49730
                     /estimated_length=unknown
     gap             53882. .53981
                     /estimated_length=unknown
     gap             57412. .57511
                     /estimated_length=unknown
     gap             62048. .62147
                     /estimated_length=unknown
     gap             65502. .65601
                     /estimated_length=unknown
     gap             67625. .67724
                     /estimated_length=unknown
     gap             69813. .69912
                     /estimated_length=unknown
     gap             72865. .72964
                     /estimated_length=unknown
     gap             75636. .75735
                     /estimated_length=unknown
     gap             78538. .78637
                     /estimated_length=unknown
     gap             83389. .83488
                     /estimated_length=unknown
     gap             86085. .86184
                     /estimated_length=unknown
     gap             88311. .88410

Alignment Scores:
Pred. No.:              1.79e+04        Length:         90572
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AC165547 (1-90572)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      61697 CATAATCATAATCATAAC 61680

RESULT 534
AC092679/c
LOCUS       AC092679               90621 bp    DNA     linear   PRI 30-APR-2005
DEFINITION  Homo sapiens BAC clone RP11-676D22 from 2, complete sequence.
ACCESSION   AC092679 AC025339
VERSION     AC092679.2  GI:19034016
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 90621)
  AUTHORS   Cotton,M., Du,H., Elliott,G. and Doebber,A.
  TITLE     The sequence of Homo sapiens BAC clone RP11-676D22
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 90621)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3  (bases 1 to 90621)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   4  (bases 1 to 90621)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-APR-2005) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Mar 1, 2002 this sequence version replaced gi:14916263.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0676D22
                Drafting Center: WIBR
            --------------.

            NOTICE:

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR:  pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-92F20, 2000 bp overlap; the
            clone sequenced to the right is RP11-811O14.  Actual start of this
            clone is at base position 34064 of RP11-92F20; actual end is at
            base position 90621 of RP11-676D22.

            The sequence of AC025339 has been incorporated into AC092679.
FEATURES                Location/Qualifiers
     source             1. .90621
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosome="2"
                        /clone="RP11-676D22"
                        /clone_lib="RPCI-11"
     misc_feature       11330. .12561
                        /note="CpG island (%GC=75.8, o/e=0.96, #CpGs=158)"
     gene               complement(11734. .12115)
                        /gene="CUL3"
```

```
     mRNA            complement(11734. .12115)
                     /gene="CUL3"
     CDS             complement(<11734. .11799)
                     /gene="CUL3"
                     /note="Homo sapiens cullin 3 (CUL3), mRNA.; H_NH0676D22.1
                     This gene was based on gi(4503164)
                     Continues as H_NH0092F20.1"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="AAY24157.1"
                     /db_xref="GI:62988770"
                     /translation="MSNLSKGTGSRKDTKMRIRAFP"
     misc_feature    30477. .30867
                     /note="CpG_island (%GC=73.4, o/e=0.67, #CpGs=37)"
     misc_feature    82095. .82322
                     /note="CpG_island (%GC=73.2, o/e=0.50, #CpGs=19)"
ORIGIN

Alignment Scores:
Pred. No.:                1.79e+04        Length:          90621
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       8               Gaps:            0

US-09-858-332G-28 (1-6) x AC092679 (1-90621)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      59339 CATAATCATAATCATAAT 59322

RESULT 535
AY102285/c
LOCUS       AY102285               90756 bp    DNA     linear   PRI 29-JAN-2003
DEFINITION  Homo sapiens RTN4 (RTN4) gene, complete sequence, alternatively
            spliced.
ACCESSION   AY102285
VERSION     AY102285.1  GI:23379818
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 90756)
  AUTHORS   Oertle,T., Huber,C., van der Putten,H. and Schwab,M.E.
  TITLE     Genomic Structure and Functional Characterisation of the Promoters
            of Human and Mouse nogo/rtn4
  JOURNAL   J. Mol. Biol. 325 (2), 299-323 (2003)
   PUBMED   12488097
REFERENCE   2  (bases 1 to 90756)
  AUTHORS   Oertle,T. and Schwab,M.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAY-2002) Brain Research Institute, University of
            Zurich and ETH Zurich, Winterthurerstr.190, Zuerich 8057,
            Switzerland
REFERENCE   3  (bases 1 to 90756)
  AUTHORS   Van der Putten,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAY-2002) Nervous System Research, Novartis Pharma
            Inc., Basel, Switzerland
FEATURES             Location/Qualifiers
     source          1. .90756
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2p16"
     repeat_region   114. .199
                     /note="L2"
                     /rpt_family="LINE/L2"
                     /rpt_type=dispersed
     repeat_region   complement(416. .429)
                     /note="L1MA4A"
                     /rpt_family="LINE"
                     /rpt_type=dispersed
     repeat_region   complement(430. .659)
                     /note="L1PB4"
                     /rpt_family="LINE"
                     /rpt_type=dispersed
     repeat_region   664. .727
                     /note="L3"
                     /rpt_family="LINE"
                     /rpt_type=dispersed
     repeat_region   complement(827. .1135)
                     /note="AluSq"
                     /rpt_family="SINE/Alu"
                     /rpt_type=dispersed
     repeat_region   complement(2136. .2485)
                     /note="L1MC2"
                     /rpt_family="LINE"
                     /rpt_type=dispersed
     repeat_region   complement(2606. .2906)
                     /note="MER33"
                     /rpt_family="MER1-type"
                     /rpt_type=dispersed
     gene            3715. .82382
                     /gene="RTN4"
                     /note="synonym: NOGO"
     mRNA            join(3715. .3827,26322. .26378,27057. .29456,66846. .67053,
                     71892. .72030,79807. .79876,80704. .80750,80950. .81008,
                     81373. .82381)
                     /gene="RTN4"
                     /product="RNT4"
                     /note="alternatively spliced; transcript variant Aa;
                     RTN4-Aa"
     exon            3715. .3827
                     /gene="RTN4"
                     /number=1Aa
     mRNA            join(3946. .4116,26322. .26378,27057. .29456,66846. .67053,
                     71892. .72030,79807. .79876,80704. .80750,80950. .81008,
                     81373. .82381)
                     /gene="RTN4"
                     /product="RNT4"
                     /note="alternatively spliced; transcript variant Ab;
                     RTN4-Ab"
     exon            3946. .4116
                     /gene="RTN4"
                     /number=1Ab
     mRNA            join(3999. .4799,26322. .26378,27057. .29456,66846. .67053,
                     71892. .72030,79807. .79876,80704. .80750,80950. .81008,
                     81373. .82381)
                     /gene="RTN4"
                     /product="RTN4"
                     /note="alternatively spliced; transcript variant A;
                     RTN4-A"
     mRNA            join(3999. .4799,26322. .26378,66846. .67053,71892. .72030,
                     79807. .79876,80704. .80750,80950. .81008,81373. .82381)
                     /gene="RTN4"
                     /product="RTN4"
                     /note="alternatively spliced; transcript variant B2;
                     RTN4-B2"
     mRNA            join(3999. .4799,66846. .67053,71892. .72030,79807. .79876,
                     80704. .80750,80950. .81008,81373. .82381)
                     /gene="RTN4"
                     /product="RTN4"
                     /note="alternatively spliced; transcript variant B1;
                     RTN4-B1"
     exon            3999. .4799
                     /gene="RTN4"
                     /number=1A
     CDS             join(4244. .4799,26322. .26378,27057. .29456,66846. .67053,
                     71892. .72030,79807. .79876,80704. .80750,80950. .81008,
                     81373. .81415)
                     /gene="RTN4"
```

```
                /note="RTN4-A; alternatively spliced"
                /codon_start=1
                /product="RTN4"
                /protein_id="AAM64243.1"
                /db_xref="GI:23379822"
                /translation="MEDLDQSPLVSSSDSPPRPQPAFKYQFVREPEDEEEEEEEEEED
                EDEDLEELEVLERKPAAGLSAAPVPTAPAAGAPLMDFGNDFVPPAPRGPLPAAPPVAP
                ERQPSWDPSPVSSTVPAPSPLSAAAVSPSKLPEDDEPPARPPPPPPASVSPQAEPVWT
                PPAPAPAAPPSTPAAPKRRGSSGSVDETLFALPAASEPVIRSSAENMDLKEQPGNTIS
                AGQEDFPSVLLETAASLPSLSPLSAASFKEHEYLGNLSTVLPTEGTLQENVSEASKEV
                SEKAKTLLIDRDLTEFSELEYSEMGSSFSVSPKAESAVIVANPREEIIVKNKDEEEKL
                VSNNILHNQQELPTALTKLVKEDEVVSSEKAKDSFNEKRVAVEAPMREEYADFKPFER
                VWEVKDSKEDSDMLAAGGKIESNLESKVDKKCFADSLEQTNHEKDSESSNDDTSFPST
                PEGIKDRSGAYITCAPFNPAATESIATNIFPLLGDPTSENKTDEKKIEEKKAQIVTEK
                NTSTKTSNPFLVAAQDSETDYVTTDNLTKVTEEVVANMPEGLTPDLVQEACESELNEV
                TGTKIAYETKMDLVQTSEVMQESLYPAAQLCPSFEESEATPSPVLPDIVMEAPLNSAV
                PSAGASVIQPSSSPLEASSVNYESIKHEPENPPPYEEAMSVSLKKVSGIKEEIKEPEN
                INAALQETEAPYISIACDLIKETKLSAEPAPDFSDYSEMAKVEQPVPDHSELVEDSSP
                DSEPVDLFSDDSIPDVPQKQDETVMLVKESLTETSFESMIEYENKEKLSALPPEGGKP
                YLESFKLSLDNTKDTLLPDEVSTLSKKEKIPLQMEELSTAVYSNDDLFISKEAQIRET
                ETFSDSSPIEIIDEFPTLISSKTDSFSKLAREYTDLEVSHKSEIANAPDGAGSLPCTE
                LPHDLSLKNIQPKVEEKISFSDDFSKNGSATSKVLLLPPDVSALATQAEIESIVKPKV
                LVKEAEKKLPSDTEKEDRSPSAIFSAELSKTSVVDLLYWRDIKKTGVVFGASLFLLLS
                LTVFSIVSVTAYIALALLSVTISFRIYKGVIQAIQKSDEGHPFRAYLESEVAISEELV
                QKYSNSALGHVNCTIKELRRLFLVDDLVDSLKFAVLMWVFTYVGALFNGLTLLILALI
                SLFSVPVIYERHQAQIDHYLGLANKNVKDAMAKIQAKIPGLKRKAE"
CDS             join(4244. .4799,26322. .26378,66846. .67053,71892. .72030,
                79807. .79876,80704. .80750,80950. .81008,81373. .81415)
                /gene="RTN4"
                /note="RTN4-B2; alternatively spliced"
                /codon_start=1
                /product="RTN4"
                /protein_id="AAM64242.1"
                /db_xref="GI:23379821"
                /translation="MEDLDQSPLVSSSDSPPRPQPAFKYQFVREPEDEEEEEEEEEED
                EDEDLEELEVLERKPAAGLSAAPVPTAPAAGAPLMDFGNDFVPPAPRGPLPAAPPVAP
                ERQPSWDPSPVSSTVPAPSPLSAAAVSPSKLPEDDEPPARPPPPPPASVSPQAEPVWT
                PPAPAPAAPPSTPAAPKRRGSSGSVDETLFALPAASEPVIRSSAVVDLLYWRDIKKTG
                VVFGASLFLLLSLTVFSIVSVTAYIALALLSVTISFRIYKGVIQAIQKSDEGHPFRAY
                LESEVAISEELVQKYSNSALGHVNCTIKELRRLFLVDDLVDSLKFAVLMWVFTYVGAL
                FNGLTLLILALISLFSVPVIYERHQAQIDHYLGLANKNVKDAMAKIQAKIPGLKRKAE
                "
CDS             join(4244. .4799,66846. .67053,71892. .72030,79807. .79876,
                80704. .80750,80950. .81008,81373. .81415)
                /gene="RTN4"
                /note="RTN4-B1; alternatively spliced"
                /codon_start=1
                /product="RTN4"
                /protein_id="AAM64241.1"
                /db_xref="GI:23379820"
                /translation="MEDLDQSPLVSSSDSPPRPQPAFKYQFVREPEDEEEEEEEEEED
                EDEDLEELEVLERKPAAGLSAAPVPTAPAAGAPLMDFGNDFVPPAPRGPLPAAPPVAP
                ERQPSWDPSPVSSTVPAPSPLSAAAVSPSKLPEDDEPPARPPPPPPASVSPQAEPVWT
                PPAPAPAAPPSTPAAPKRRGSSGSVVVDLLYWRDIKKTGVVFGASLFLLLSLTVFSIV
                SVTAYIALALLSVTISFRIYKGVIQAIQKSDEGHPFRAYLESEVAISEELVQKYSNSA
                LGHVNCTIKELRRLFLVDDLVDSLKFAVLMWVFTYVGALFNGLTLLILALISLFSVPV
                IYERHQAQIDHYLGLANKNVKDAMAKIQAKIPGLKRKAE"
repeat_region   4337. .4380
                /rpt_type=tandem
                /rpt_unit="gga"
repeat_region   4419. .4849
                /note="L1M2-like"
                /rpt_family="LINE-like"
                /rpt_type=dispersed
mRNA            join(4853. .4986,26322. .26378,27057. .29456,66846. .67053,
                71892. .72030,79807. .79876,80704. .80750,80950. .81008,
                81373. .82381)
                /gene="RTN4"
                /product="RNT4"
                /note="alternatively spliced; transcript variant D;
                RTN4-D"
exon            4853. .4986
                /gene="RTN4"
                /number=1D
mRNA            join(5346. .5465,26322. .26378,27057. .29456,66846. .67053,
                71892. .72030,79807. .79876,80704. .80750,80950. .81008,
                81373. .82381)
                /gene="RTN4"
                /product="RNT4"
                /note="alternatively spliced; transcript variant E;
                RTN4-E"
exon            5346. .5465
                /gene="RTN4"
                /number=1E
repeat_region   5864. .5971
                /note="Tigger3"
                /rpt_family="MER2-type"
                /rpt_type=dispersed

Alignment Scores:
Pred. No.:              1.8e+04         Length:         90756
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     8               Gaps:           0

US-09-858-332G-28 (1-6) x AY102285 (1-90756)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       7759 CACAACCATAATCACAAT 7742

RESULT 536
AC153083/c
LOCUS       AC153083               91096 bp    DNA     linear   HTG 22-JUN-2005
DEFINITION  Ornithorhynchus anatinus clone OA_Bb-241H13, WORKING DRAFT
            SEQUENCE, 9 unordered pieces.
ACCESSION   AC153083
VERSION     AC153083.2  GI:68131668
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Ornithorhynchus anatinus (platypus)
  ORGANISM  Ornithorhynchus anatinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Monotremata; Ornithorhynchidae; Ornithorhynchus.
REFERENCE   1  (bases 1 to 91096)
  AUTHORS   Antonellis,A., Ayele,K., Bass,D., Benjamin,B., Bera,J.,
            Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G.,
            Coleman,H., Engle,J., Fuksenko,T., Gestole,M., Greene,A., Guan,X.,
            Gupta,J., Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Hu,P.,
            Hunter,G., Hurle,B., Idol,J.R., Kwong,P., Laric,P., Larson,S.,
            Lee-Lin,S.-Q., Legaspi,R., Madden,M., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Masiello,C., Maskeri,B., McDowell,J., Mojidi,H.A.,
            Mullikin,J.C., Oestreicher,J.S., Park,M., Portnoy,M.E., Prasad,A.,
            Puri,O., Reddix-Dugue,N., Sante,A., Schandler,K., Schueler,M.G.,
            Sison,C., Stantripop,S., Stephen,E., Taye,A., Thomas,J.W.,
            Thomas,P.J., Tsipouri,V., Ung,L., Vogt,J.L., Wetherby,K.D.,
            Withers,T.R., Young,A. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 91096)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-DEC-2004) NIH Intramural Sequencing Center, 5625
            Fishers Lane, Rockville, MD 20852, USA
REFERENCE   3  (bases 1 to 91096)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JUN-2005) NIH Intramural Sequencing Center, 5625
            Fishers Lane, Rockville, MD 20852, USA
COMMENT     On Jun 22, 2005 this sequence version replaced gi:56159017.
            -------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
```

```
            ---------- Project Information
                 Center project name: geq
                 Center clone name: 241H13
            ---------- Summary Statistics
                 Sequencing vector: plasmid; n/a; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 85476 bases at least Q40
                 Consensus quality: 86471 bases at least Q30
                 Consensus quality: 87022 bases at least Q20
                 Insert size: 126000; agarose-fp
                 Insert size: 90296; sum-of-contigs
                 Quality coverage: 8.00x in Q20 bases; agarose-fp
                 Quality coverage: 11.16x in Q20 bases; sum-of-contigs
            ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3276: contig of 3276 bp in length
            *     3277     3376: gap of unknown length
            *     3377     6684: contig of 3308 bp in length
            *     6685     6784: gap of unknown length
            *     6785     8943: contig of 2159 bp in length
            *     8944     9043: gap of unknown length
            *     9044    11719: contig of 2676 bp in length
            *    11720    11819: gap of unknown length
            *    11820    15866: contig of 4047 bp in length
            *    15867    15966: gap of unknown length
            *    15967    34311: contig of 18345 bp in length
            *    34312    34411: gap of unknown length
            *    34412    40849: contig of 6438 bp in length
            *    40850    40949: gap of unknown length
            *    40950    59249: contig of 18300 bp in length
            *    59250    59349: gap of unknown length
            *    59350    91096: contig of 31747 bp in length.
FEATURES             Location/Qualifiers
     source          1. .91096
                     /organism="Ornithorhynchus anatinus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9258"
                     /clone="OA_Bb-241H13"
                     /clone_lib="OA_Bb"
     misc_feature    1. .3276
                     /note="assembly_fragment"
     gap             3277. .3376
                     /estimated_length=unknown
     misc_feature    3377. .6684
                     /note="assembly_fragment"
     gap             6685. .6784
                     /estimated_length=unknown
     misc_feature    6785. .8943
                     /note="assembly_fragment"
     gap             8944. .9043
                     /estimated_length=unknown
     misc_feature    9044. .11719
                     /note="assembly_fragment"
     gap             11720. .11819
                     /estimated_length=unknown
     misc_feature    11820. .15866
                     /note="assembly_fragment"
     gap             15867. .15966
                     /estimated_length=unknown
     misc_feature    15967. .34311
                     /note="assembly_fragment"
     gap             34312. .34411
                     /estimated_length=unknown
     misc_feature    34412. .40849
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     gap             40850. .40949
                     /estimated_length=unknown
     misc_feature    40950. .59249
                     /note="assembly_fragment"
     gap             59250. .59349
                     /estimated_length=unknown
     misc_feature    59350. .91096
                     /note="assembly_fragment"
ORIGIN

Alignment Scores:
Pred. No.:                1.8e+04        Length:          91096
Score:                    42.00          Matches:         6
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       14             Gaps:            0

US-09-858-332G-28 (1-6) x AC153083 (1-91096)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       1793 CATAATCATAATCATAAT 1776

RESULT 537
HSDA86H19
LOCUS       HSDA86H19               91201 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP6-86H19 on chromosome Xq25-26.2,
            complete sequence.
ACCESSION   AL049562
VERSION     AL049562.7  GI:4958821
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 91201)
  AUTHORS   Howden,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 1, 1999 this sequence version replaced gi:4741573.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            RP6-86H19 is from the library RPCI-6 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pPAC4
            ---------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            ----------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
```

```
     source          1. .91201
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="RZPD:RPCIP709H1986"
                     /db_xref="taxon:9606"
                     /map="q25-26.2"
                     /clone="RP6-86H19"
                     /clone_lib="RPCI-6"
ORIGIN

Alignment Scores:
Pred. No.:              1.81e+04        Length:         91201
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     8               Gaps:           0

US-09-858-332G-28 (1-6) x HSDA86H19 (1-91201)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      87892 CATAATCATAATCATAAT 87909

RESULT 538
AP008143/c
LOCUS       AP008143               91272 bp    DNA     linear   HTG 28-DEC-2004
DEFINITION  Lotus corniculatus var. japonicus chromosome 1 clone LjT32B13, ***
            SEQUENCING IN PROGRESS ***, 15 unordered pieces.
ACCESSION   AP008143
VERSION     AP008143.1  GI:56806450
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1
  AUTHORS   Kaneko,T., Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Structural Analysis of a Lotus japonicus Genome. XI. Sequence
            Features and Mapping of Nine hundred twenty-one TAC Clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 91272)
  AUTHORS   Sato,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-2004) Shusei Sato, Kazusa DNA Research Institute,
            Department of Plant Gene Research; 2-6-7 Kazusa-kamatari, Kisarazu,
            Chiba, 292-0818, Japan (E-mail:ssato@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/, Tel:81-438-52-3935(ex.2337),
            Fax:81-438-52-3934)
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 15 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     6536: contig of 6536 bp in length
            *     6537     6636: gap of unknown length
            *     6637     8242: contig of 1606 bp in length
            *     8243     8342: gap of unknown length
            *     8343    10157: contig of 1815 bp in length
            *    10158    10257: gap of unknown length
            *    10258    11511: contig of 1254 bp in length
            *    11512    11611: gap of unknown length
            *    11612    13520: contig of 1909 bp in length
            *    13521    13620: gap of unknown length
            *    13621    15878: contig of 2258 bp in length
            *    15879    15978: gap of unknown length
            *    15979    18104: contig of 2126 bp in length
            *    18105    18204: gap of unknown length
            *    18205    21848: contig of 3644 bp in length
            *    21849    21948: gap of unknown length
            *    21949    26506: contig of 4558 bp in length
            *    26507    26606: gap of unknown length
            *    26607    32341: contig of 5735 bp in length
            *    32342    32441: gap of unknown length
            *    32442    37453: contig of 5012 bp in length
            *    37454    37553: gap of unknown length
            *    37554    46358: contig of 8805 bp in length
            *    46359    46458: gap of unknown length
            *    46459    62278: contig of 15820 bp in length
            *    62279    62378: gap of unknown length
            *    62379    88123: contig of 25745 bp in length
            *    88124    88223: gap of unknown length
            *    88224    91272: contig of 3049 bp in length.
FEATURES             Location/Qualifiers
     source          1. .91272
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="genomic DNA"
                     /variety="japonicus"
                     /db_xref="taxon:34305"
                     /chromosome="1"
                     /clone="LjT32B13"
                     /clone_lib="LjT library"
                     /note="TAC clone:TM1542, synonym:Lotus japonicus"
     gap             6537. .6636
                     /estimated_length=unknown
     gap             8243. .8342
                     /estimated_length=unknown
     gap             10158. .10257
                     /estimated_length=unknown
     gap             11512. .11611
                     /estimated_length=unknown
     gap             13521. .13620
                     /estimated_length=unknown
     gap             15879. .15978
                     /estimated_length=unknown
     gap             18105. .18204
                     /estimated_length=unknown
     gap             21849. .21948
                     /estimated_length=unknown
     gap             26507. .26606
                     /estimated_length=unknown
     gap             32342. .32441
                     /estimated_length=unknown
     gap             37454. .37553
                     /estimated_length=unknown
     gap             46359. .46458
                     /estimated_length=unknown
     gap             62279. .62378
                     /estimated_length=unknown
     gap             88124. .88223
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              1.81e+04        Length:         91272
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AP008143 (1-91272)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      80922 CACAATCACAATCATAAT 80905

RESULT 539
AC130798/c
LOCUS       AC130798               92189 bp    DNA     linear   PLN 08-AUG-2004
DEFINITION  Medicago truncatula clone mth2-27f17, complete sequence.
```

ACCESSION AC130798
VERSION AC130798.15 GI:51011169
KEYWORDS HTG.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE 1 (bases 1 to 92189)
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A.
TITLE Medicago truncatula BAC Clone mth2-27f17
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 92189)
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (14-AUG-2002) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
REFERENCE 3 (bases 1 to 92189)
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2003) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
REFERENCE 4 (bases 1 to 92189)
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (31-MAY-2003) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
REFERENCE 5 (bases 1 to 92189)
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (05-JUN-2003) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
REFERENCE 6 (bases 1 to 92189)
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (09-MAR-2004) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
REFERENCE 7 (bases 1 to 92189)
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (22-JUN-2004) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
REFERENCE 8 (bases 1 to 92189)
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (22-JUL-2004) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
REFERENCE 9 (bases 1 to 92189)
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-2004) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
REFERENCE 10 (bases 1 to 92189)
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (06-AUG-2004) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
REFERENCE 11 (bases 1 to 92189)
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (08-AUG-2004) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
COMMENT On Aug 6, 2004 this sequence version replaced gi:49034921.
---------- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
----------.
FEATURES Location/Qualifiers
source 1. .92189
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/db_xref="taxon:3880"
/clone="mth2-27f17"
/clone_lib="Medicago truncatula BAC library H2"
ORIGIN

Alignment Scores:

Pred. No.:	1.82e+04	Length:	92189
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	15	Gaps:	0

US-09-858-332G-28 (1-6) x AC130798 (1-92189)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      33008 CATAACCATAACCATAAC 32991
```

RESULT 540
BX842667
LOCUS BX842667 92969 bp DNA linear ROD 12-DEC-2003
DEFINITION Mouse DNA sequence from clone RP24-319D9 on chromosome 11, complete sequence.
ACCESSION BX842667
VERSION BX842667.2 GI:39840710
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 92969)
AUTHORS Matthews,L.
TITLE Direct Submission
JOURNAL Submitted (12-DEC-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Dec 12, 2003 this sequence version replaced gi:38567348.
Sequence from the Mouse Genome Sequencing Consortium whole genome shotgun may have been used to confirm this sequence. Sequence data from the whole genome shotgun alone has only been used where it has a phred quality of at least 30.
---------- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the

```
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP24-319D9 is
            from a Male (C57BL/6J) mouse BAC Library VECTOR: pTARBAC1.
FEATURES              Location/Qualifiers
     source           1. .92969
                      /organism="Mus musculus"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:10090"
                      /chromosome="11"
                      /clone="RP24-319D9"
                      /clone_lib="RPCI-24"
ORIGIN

Alignment Scores:
Pred. No.:              1.84e+04        Length:         92969
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     9               Gaps:           0

US-09-858-332G-28 (1-6) x BX842667 (1-92969)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      40132 CACAACCACAACCACAAC 40149

RESULT 541
AC140159/c
LOCUS       AC140159               92994 bp    DNA     linear   HTG 03-JUN-2005
DEFINITION  Latimeria menadoensis clone vmrc1-4a18, WORKING DRAFT SEQUENCE, 3
            ordered pieces.
ACCESSION   AC140159
VERSION     AC140159.15  GI:66912372
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Latimeria menadoensis (Menado coelacanth)
  ORGANISM  Latimeria menadoensis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Coelacanthiformes; Coelacanthidae; Latimeria.
REFERENCE   1  (bases 1 to 92994)
  AUTHORS   Lau,C.C.Y., Amemiya,C. and Roe,B.A.
  TITLE     Latimeria menadoensis BAC Clone vmrc1-4a18
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 92994)
  AUTHORS   Lau,C.C.Y., Amemiya,C. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 92994)
  AUTHORS   Lau,C.C.Y., Amemiya,C. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUN-2005) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Jun 3, 2005 this sequence version replaced gi:61656342.
            -------------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ----------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    56160: contig of 56160 bp in length
            *    56161    56260: gap of unknown length
            *    56261    70894: contig of 14634 bp in length
            *    70895    70994: gap of unknown length
            *    70995    92994: contig of 22000 bp in length.
FEATURES              Location/Qualifiers
     source           1. .92994
                      /organism="Latimeria menadoensis"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:106881"
                      /clone="vmrc1-4a18"
                      /clone_lib="Latimeria menadoensis BAC Library 1"
     gap              56161. .56260
                      /estimated_length=unknown
     gap              70895. .70994
                      /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              1.84e+04        Length:         92994
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AC140159 (1-92994)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      49179 CACAACCACAACCACAAC 49162

RESULT 542
AC138752/c
LOCUS       AC138752               93305 bp    DNA     linear   HTG 16-JAN-2003
DEFINITION  Homo sapiens chromosome 17 clone RP11-1022D1 map 17, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC138752
VERSION     AC138752.1  GI:27764725
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 93305)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-1022D1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 93305)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
```

Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (16-JAN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
---------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
---------- Project Information
Center project name: L29179
Center clone name: 1022_D_1
----------.
* NOTE: This record contains 88 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

*	1	980	contig of 980 bp in length
*	981	1080	gap of 100 bp
*	1081	2070	contig of 990 bp in length
*	2071	2170	gap of 100 bp
*	2171	3170	contig of 1000 bp in length
*	3171	3270	gap of 100 bp
*	3271	4199	contig of 929 bp in length
*	4200	4299	gap of 100 bp
*	4300	5255	contig of 956 bp in length
*	5256	5355	gap of 100 bp
*	5356	6327	contig of 972 bp in length
*	6328	6427	gap of 100 bp
*	6428	7427	contig of 1000 bp in length
*	7428	7527	gap of 100 bp
*	7528	8492	contig of 965 bp in length
*	8493	8592	gap of 100 bp
*	8593	9533	contig of 941 bp in length
*	9534	9633	gap of 100 bp
*	9634	10574	contig of 941 bp in length
*	10575	10674	gap of 100 bp
*	10675	11702	contig of 1028 bp in length
*	11703	11802	gap of 100 bp
*	11803	12737	contig of 935 bp in length
*	12738	12837	gap of 100 bp
*	12838	13753	contig of 916 bp in length
*	13754	13853	gap of 100 bp
*	13854	14740	contig of 887 bp in length
*	14741	14840	gap of 100 bp
*	14841	15830	contig of 990 bp in length
*	15831	15930	gap of 100 bp
*	15931	16896	contig of 966 bp in length
*	16897	16996	gap of 100 bp
*	16997	18000	contig of 1004 bp in length
*	18001	18100	gap of 100 bp
*	18101	19095	contig of 995 bp in length
*	19096	19195	gap of 100 bp
*	19196	20150	contig of 955 bp in length
*	20151	20250	gap of 100 bp
*	20251	21195	contig of 945 bp in length
*	21196	21295	gap of 100 bp
*	21296	22256	contig of 961 bp in length
*	22257	22356	gap of 100 bp
*	22357	23344	contig of 988 bp in length
*	23345	23444	gap of 100 bp
*	23445	24431	contig of 987 bp in length
*	24432	24531	gap of 100 bp
*	24532	25509	contig of 978 bp in length
*	25510	25609	gap of 100 bp
*	25610	26553	contig of 944 bp in length
*	26554	26653	gap of 100 bp
*	26654	27652	contig of 999 bp in length
*	27653	27752	gap of 100 bp
*	27753	28731	contig of 979 bp in length
*	28732	28831	gap of 100 bp
*	28832	29855	contig of 1024 bp in length
*	29856	29955	gap of 100 bp
*	29956	30972	contig of 1017 bp in length
*	30973	31072	gap of 100 bp
*	31073	32099	contig of 1027 bp in length
*	32100	32199	gap of 100 bp
*	32200	33181	contig of 982 bp in length
*	33182	33281	gap of 100 bp
*	33282	34184	contig of 903 bp in length
*	34185	34284	gap of 100 bp
*	34285	35285	contig of 1001 bp in length
*	35286	35385	gap of 100 bp
*	35386	36352	contig of 967 bp in length
*	36353	36452	gap of 100 bp
*	36453	37356	contig of 904 bp in length
*	37357	37456	gap of 100 bp
*	37457	38433	contig of 977 bp in length
*	38434	38533	gap of 100 bp
*	38534	39566	contig of 1033 bp in length
*	39567	39666	gap of 100 bp
*	39667	40683	contig of 1017 bp in length
*	40684	40783	gap of 100 bp
*	40784	41753	contig of 970 bp in length
*	41754	41853	gap of 100 bp
*	41854	42846	contig of 993 bp in length
*	42847	42946	gap of 100 bp
*	42947	43928	contig of 982 bp in length
*	43929	44028	gap of 100 bp
*	44029	45008	contig of 980 bp in length
*	45009	45108	gap of 100 bp
*	45109	46080	contig of 972 bp in length
*	46081	46180	gap of 100 bp
*	46181	47141	contig of 961 bp in length
*	47142	47241	gap of 100 bp
*	47242	48168	contig of 927 bp in length
*	48169	48268	gap of 100 bp
*	48269	49219	contig of 951 bp in length
*	49220	49319	gap of 100 bp
*	49320	50247	contig of 928 bp in length
*	50248	50347	gap of 100 bp
*	50348	51273	contig of 926 bp in length
*	51274	51373	gap of 100 bp
*	51374	52361	contig of 988 bp in length
*	52362	52461	gap of 100 bp
*	52462	53447	contig of 986 bp in length
*	53448	53547	gap of 100 bp
*	53548	54542	contig of 995 bp in length
*	54543	54642	gap of 100 bp
*	54643	55605	contig of 963 bp in length
*	55606	55705	gap of 100 bp
*	55706	56695	contig of 990 bp in length
*	56696	56795	gap of 100 bp
*	56796	57759	contig of 964 bp in length
*	57760	57859	gap of 100 bp
*	57860	58846	contig of 987 bp in length
*	58847	58946	gap of 100 bp
*	58947	59895	contig of 949 bp in length
*	59896	59995	gap of 100 bp
*	59996	60962	contig of 967 bp in length
*	60963	61062	gap of 100 bp

```
            *    61063    61872: contig of 810 bp in length
            *    61873    61972: gap of 100 bp
            *    61973    62912: contig of 940 bp in length
            *    62913    63012: gap of 100 bp
            *    63013    63991: contig of 979 bp in length
            *    63992    64091: gap of 100 bp
            *    64092    65065: contig of 974 bp in length
            *    65066    65165: gap of 100 bp
            *    65166    66146: contig of 981 bp in length
            *    66147    66246: gap of 100 bp
            *    66247    67234: contig of 988 bp in length
            *    67235    67334: gap of 100 bp
            *    67335    68312: contig of 978 bp in length
            *    68313    68412: gap of 100 bp
            *    68413    69326: contig of 914 bp in length
            *    69327    69426: gap of 100 bp
            *    69427    70361: contig of 935 bp in length
            *    70362    70461: gap of 100 bp
            *    70462    71441: contig of 980 bp in length
            *    71442    71541: gap of 100 bp
            *    71542    72403: contig of 862 bp in length
            *    72404    72503: gap of 100 bp
            *    72504    73490: contig of 987 bp in length
            *    73491    73590: gap of 100 bp
            *    73591    74587: contig of 997 bp in length
            *    74588    74687: gap of 100 bp

Alignment Scores:
Pred. No.:              1.85e+04        Length:         93305
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AC138752 (1-93305)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      46657 CATAATCATAATCATAAT 46640

RESULT 543
AC136849
LOCUS       AC136849               94175 bp    DNA     linear   HTG 09-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-320N14, *** SEQUENCING IN PROGRESS
            ***, 49 unordered pieces.
ACCESSION   AC136849
VERSION     AC136849.1  GI:24818530
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 94175)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., Mcleod,M., Mcneill,T., Meenen,E., Milosavljevic,A.,
            Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K.,
            Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D.,
            Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G.,
            Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H.,
            Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A.,
            Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E.,
            Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y.,
            Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A.,
            Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S.,
            Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A.,
            Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E.,
            Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A.,
            Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S.,
            Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D.,
            Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J.,
            Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R.,
            Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S.,
            Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X.,
            Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R.,
            Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 94175)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: KCTF
                Center clone name: CH230-320N14
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 50389 bases at least Q40
                Consensus quality: 52929 bases at least Q30
                Consensus quality: 54643 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 49 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1001: contig of 1001 bp in length
            *     1002     1101: gap of unknown length
            *     1102     2558: contig of 1457 bp in length
            *     2559     2658: gap of unknown length
            *     2659     3986: contig of 1328 bp in length
            *     3987     4086: gap of unknown length
            *     4087     5311: contig of 1225 bp in length
            *     5312     5411: gap of unknown length
            *     5412     6537: contig of 1126 bp in length
```

```
*     6538     6637: gap of unknown length
*     6638     8158: contig of 1521 bp in length
*     8159     8258: gap of unknown length
*     8259     9372: contig of 1114 bp in length
*     9373     9472: gap of unknown length
*     9473    10777: contig of 1305 bp in length
*    10778    10877: gap of unknown length
*    10878    11970: contig of 1093 bp in length
*    11971    12070: gap of unknown length
*    12071    13224: contig of 1154 bp in length
*    13225    13324: gap of unknown length
*    13325    14359: contig of 1035 bp in length
*    14360    14459: gap of unknown length
*    14460    15466: contig of 1007 bp in length
*    15467    15566: gap of unknown length
*    15567    16870: contig of 1304 bp in length
*    16871    16970: gap of unknown length
*    16971    18124: contig of 1154 bp in length
*    18125    18224: gap of unknown length
*    18225    19711: contig of 1487 bp in length
*    19712    19811: gap of unknown length
*    19812    21225: contig of 1414 bp in length
*    21226    21325: gap of unknown length
*    21326    22577: contig of 1252 bp in length
*    22578    22677: gap of unknown length
*    22678    24599: contig of 1922 bp in length
*    24600    24699: gap of unknown length
*    24700    26088: contig of 1389 bp in length
*    26089    26188: gap of unknown length
*    26189    27786: contig of 1598 bp in length
*    27787    27886: gap of unknown length
*    27887    29594: contig of 1708 bp in length
*    29595    29694: gap of unknown length
*    29695    31558: contig of 1864 bp in length
*    31559    31658: gap of unknown length
*    31659    33946: contig of 2288 bp in length
*    33947    34046: gap of unknown length
*    34047    35519: contig of 1473 bp in length
*    35520    35619: gap of unknown length
*    35620    37281: contig of 1662 bp in length
*    37282    37381: gap of unknown length
*    37382    39144: contig of 1763 bp in length
*    39145    39244: gap of unknown length
*    39245    40250: contig of 1006 bp in length
*    40251    40350: gap of unknown length
*    40351    41999: contig of 1649 bp in length
*    42000    42099: gap of unknown length
*    42100    44764: contig of 2665 bp in length
*    44765    44864: gap of unknown length
*    44865    47771: contig of 2907 bp in length
*    47772    47871: gap of unknown length
*    47872    49509: contig of 1638 bp in length
*    49510    49609: gap of unknown length
*    49610    51569: contig of 1960 bp in length
*    51570    51669: gap of unknown length
*    51670    53560: contig of 1891 bp in length
*    53561    53660: gap of unknown length
*    53661    55808: contig of 2148 bp in length
*    55809    55908: gap of unknown length
*    55909    58413: contig of 2505 bp in length
*    58414    58513: gap of unknown length
*    58514    61181: contig of 2668 bp in length
*    61182    61281: gap of unknown length
*    61282    63624: contig of 2343 bp in length
*    63625    63724: gap of unknown length
*    63725    65847: contig of 2123 bp in length
*    65848    65947: gap of unknown length
*    65948    68486: contig of 2539 bp in length
*    68487    68586: gap of unknown length
*    68587    70657: contig of 2071 bp in length
*    70658    70757: gap of unknown length
*    70758    72514: contig of 1757 bp in length
*    72515    72614: gap of unknown length
*    72615    74307: contig of 1693 bp in length
*    74308    74407: gap of unknown length
*    74408    76928: contig of 2521 bp in length
*    76929    77028: gap of unknown length
*    77029    79099: contig of 2071 bp in length
*    79100    79199: gap of unknown length
*    79200    82355: contig of 3156 bp in length
*    82356    82455: gap of unknown length
*    82456    84627: contig of 2172 bp in length
*    84628    84727: gap of unknown length
*    84728    86444: contig of 1717 bp in length
*    86445    86544: gap of unknown length
*    86545    90882: contig of 4338 bp in length
*    90883    90982: gap of unknown length
*    90983    94175: contig of 3193 bp in length.
FEATURES             Location/Qualifiers
     source          1. .94175
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-320N14"
     gap             1002. .1101
                     /estimated_length=unknown
     gap             2559. .2658

Alignment Scores:
Pred. No.:                1.86e+04        Length:          94175
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       14              Gaps:            0

US-09-858-332G-28 (1-6) x AC136849 (1-94175)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      93348 CACAATCACAATCACAAT 93365

RESULT 544
AP007998
LOCUS       AP007998               94857 bp    DNA     linear   HTG 28-DEC-2004
DEFINITION  Lotus corniculatus var. japonicus chromosome 6 clone LjT40C18, ***
            SEQUENCING IN PROGRESS ***, 27 unordered pieces.
ACCESSION   AP007998
VERSION     AP007998.1  GI:56806306
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1
  AUTHORS   Kaneko,T., Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Structural Analysis of a Lotus japonicus Genome. XI. Sequence
            Features and Mapping of Nine hundred twenty-one TAC Clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 94857)
  AUTHORS   Sato,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-2004) Shusei Sato, Kazusa DNA Research Institute,
            Department of Plant Gene Research; 2-6-7 Kazusa-kamatari, Kisarazu,
            Chiba, 292-0818, Japan (E-mail:ssato@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/, Tel:81-438-52-3935(ex.2337),
            Fax:81-438-52-3934)
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 27 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
```

```
          * be preserved.
          *        1      645: contig of 645 bp in length
          *      646      745: gap of unknown length
          *      746     1775: contig of 1030 bp in length
          *     1776     1875: gap of unknown length
          *     1876     2676: contig of 801 bp in length
          *     2677     2776: gap of unknown length
          *     2777     3592: contig of 816 bp in length
          *     3593     3692: gap of unknown length
          *     3693     4353: contig of 661 bp in length
          *     4354     4453: gap of unknown length
          *     4454     5822: contig of 1369 bp in length
          *     5823     5922: gap of unknown length
          *     5923     6910: contig of 988 bp in length
          *     6911     7010: gap of unknown length
          *     7011     8108: contig of 1098 bp in length
          *     8109     8208: gap of unknown length
          *     8209     9736: contig of 1528 bp in length
          *     9737     9836: gap of unknown length
          *     9837    10751: contig of 915 bp in length
          *    10752    10851: gap of unknown length
          *    10852    12316: contig of 1465 bp in length
          *    12317    12416: gap of unknown length
          *    12417    14553: contig of 2137 bp in length
          *    14554    14653: gap of unknown length
          *    14654    17471: contig of 2818 bp in length
          *    17472    17571: gap of unknown length
          *    17572    20671: contig of 3100 bp in length
          *    20672    20771: gap of unknown length
          *    20772    23514: contig of 2743 bp in length
          *    23515    23614: gap of unknown length
          *    23615    26309: contig of 2695 bp in length
          *    26310    26409: gap of unknown length
          *    26410    31468: contig of 5059 bp in length
          *    31469    31568: gap of unknown length
          *    31569    36475: contig of 4907 bp in length
          *    36476    36575: gap of unknown length
          *    36576    39200: contig of 2625 bp in length
          *    39201    39300: gap of unknown length
          *    39301    44090: contig of 4790 bp in length
          *    44091    44190: gap of unknown length
          *    44191    49291: contig of 5101 bp in length
          *    49292    49391: gap of unknown length
          *    49392    55401: contig of 6010 bp in length
          *    55402    55501: gap of unknown length
          *    55502    61280: contig of 5779 bp in length
          *    61281    61380: gap of unknown length
          *    61381    68375: contig of 6995 bp in length
          *    68376    68475: gap of unknown length
          *    68476    76976: contig of 8501 bp in length
          *    76977    77076: gap of unknown length
          *    77077    93856: contig of 16780 bp in length
          *    93857    93956: gap of unknown length
          *    93957    94857: contig of 901 bp in length.
FEATURES             Location/Qualifiers
     source          1. .94857
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="genomic DNA"
                     /variety="japonicus"
                     /db_xref="taxon:34305"
                     /chromosome="6"
                     /clone="LjT40C18"
                     /clone_lib="LjT library"
                     /note="TAC clone:TM1332, synonym:Lotus japonicus"
     gap             646. .745
                     /estimated_length=unknown
     gap             1776. .1875
                     /estimated_length=unknown
     gap             2677. .2776
                     /estimated_length=unknown
     gap             3593. .3692
                     /estimated_length=unknown
     gap             4354. .4453
                     /estimated_length=unknown
     gap             5823. .5922
                     /estimated_length=unknown
     gap             6911. .7010
                     /estimated_length=unknown
     gap             8109. .8208
                     /estimated_length=unknown
     gap             9737. .9836
                     /estimated_length=unknown
     gap             10752. .10851
                     /estimated_length=unknown
     gap             12317. .12416
                     /estimated_length=unknown
     gap             14554. .14653
                     /estimated_length=unknown
     gap             17472. .17571
                     /estimated_length=unknown
     gap             20672. .20771
                     /estimated_length=unknown
     gap             23515. .23614
                     /estimated_length=unknown
     gap             26310. .26409
                     /estimated_length=unknown
     gap             31469. .31568
                     /estimated_length=unknown
     gap             36476. .36575
                     /estimated_length=unknown
     gap             39201. .39300
                     /estimated_length=unknown
     gap             44091. .44190
                     /estimated_length=unknown
     gap             49292. .49391
                     /estimated_length=unknown
     gap             55402. .55501
                     /estimated_length=unknown
     gap             61281. .61380
                     /estimated_length=unknown
     gap             68376. .68475
                     /estimated_length=unknown
     gap             76977. .77076
                     /estimated_length=unknown
     gap             93857. .93956
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:                1.88e+04        Length:          94857
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       14              Gaps:            0

US-09-858-332G-28 (1-6) x AP007998 (1-94857)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      25011 CATAACCACAATCATAAT 25028

RESULT 545
AC020342/c
LOCUS       AC020342               94875 bp    DNA     linear   HTG 03-JAN-2000
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC020342
VERSION     AC020342.1  GI:6664555
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 94875)
  AUTHORS   Adams,M. and Venter,J.C.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10212841 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .94875
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
ORIGIN

Alignment Scores:
Pred. No.:                1.88e+04        Length:         94875
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       14              Gaps:           0

US-09-858-332G-28 (1-6) x AC020342 (1-94875)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      93897 CATAATCATAATCATAAT 93880

RESULT 546
AC022816
LOCUS       AC022816               95546 bp    DNA     linear   PRI 23-AUG-2000
DEFINITION  Homo sapiens chromosome 17, clone RP11-626C5, complete sequence.
ACCESSION   AC022816
VERSION     AC022816.15  GI:9886053
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 95546)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-626C5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 95546)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 95546)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
            FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
            Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
            Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G.,
            Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
            O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
            Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Aug 23, 2000 this sequence version replaced gi:9838013.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html

            Only 95546 base pairs from the middle of this clone are being
            submitted.  The remainder overlaps with either accession number
            AC006578 (WICGR project L754) or accession number AC005224 (WICGR
            project L354).
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L6104
                Center clone name: 626_C_5
            --------------.
FEATURES             Location/Qualifiers
     source          1. .95546
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-626C5"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   362. .715
                     /rpt_family="L1PA5"
     repeat_region   1671. .1724
                     /rpt_family="(TTTC)n"
     repeat_region   complement(1725. .2193)
                     /rpt_family="AluJb"
     repeat_region   2194. .2203
                     /rpt_family="(TTTC)n"
     repeat_region   2963. .3025
                     /rpt_family="L1MC4"
     repeat_region   3026. .3084
                     /rpt_family="AT_rich"
     repeat_region   3750. .3891
                     /rpt_family="AluSp"
     repeat_region   3892. .4200
                     /rpt_family="AluSp"
     repeat_region   4201. .4341
                     /rpt_family="AluSp"
     repeat_region   4342. .4671
                     /rpt_family="HAL1"
     repeat_region   5420. .5524
                     /rpt_family="L1ME"
     repeat_region   5784. .5829
                     /rpt_family="(TTTA)n"
     repeat_region   complement(5863. .6167)
                     /rpt_family="AluSp"
     repeat_region   7588. .7693
                     /rpt_family="MIR"
```

```
repeat_region   7694. .8136
                /rpt_family="L1MA9"
repeat_region   8137. .8239
                /rpt_family="MIR"
repeat_region   8690. .8748
                /rpt_family="(TTTTA)n"
repeat_region   complement(8749. .9031)
                /rpt_family="AluSg"
repeat_region   9317. .9337
                /rpt_family="AT_rich"
repeat_region   9407. .9442
                /rpt_family="(CATATA)n"
repeat_region   complement(9492. .9539)
                /rpt_family="L2"
repeat_region   complement(9764. .9842)
                /rpt_family="AluJ/FRAM"
repeat_region   9855. .9986
                /rpt_family="(ATATG)n"
repeat_region   9987. .10166
                /rpt_family="(TATG)n"
repeat_region   10211. .10768
                /rpt_family="(TA)n"
repeat_region   complement(11005. .11311)
                /rpt_family="AluYb8"
repeat_region   11857. .12137
                /rpt_family="AluSc"
repeat_region   12138. .12179
                /rpt_family="(CAAA)n"
repeat_region   complement(12393. .12603)
                /rpt_family="AluJo"
repeat_region   14202. .14374
                /rpt_family="MIR"
repeat_region   complement(14490. .14578)
                /rpt_family="MIR"
repeat_region   complement(14691. .14811)
                /rpt_family="L2"
repeat_region   15412. .15485
                /rpt_family="MIR"
repeat_region   15861. .15891
                /rpt_family="(TG)n"
repeat_region   15967. .15990
                /rpt_family="AT_rich"
repeat_region   complement(15991. .16290)
                /rpt_family="AluY"
repeat_region   16630. .16941
                /rpt_family="AluY"
repeat_region   17139. .17240
                /rpt_family="GA-rich"
repeat_region   complement(17767. .18129)
                /rpt_family="L2"
repeat_region   18155. .18282
                /rpt_family="MER20"
repeat_region   18284. .18580
                /rpt_family="AluSg"
repeat_region   18680. .18709
                /rpt_family="AT_rich"
repeat_region   complement(18710. .19178)
                /rpt_family="L1MA8"
repeat_region   complement(19270. .19954)
                /rpt_family="L2"
repeat_region   19955. .20264
                /rpt_family="AluSx"
repeat_region   complement(20265. .20335)
                /rpt_family="L2"
repeat_region   20438. .20474
                /rpt_family="(T)n"
repeat_region   complement(20866. .21219)
                /rpt_family="HAL1"
repeat_region   21230. .21303
                /rpt_family="(TAA)n"
repeat_region   complement(21768. .21868)
                /rpt_family="MIR"
repeat_region   22777. .22914
                /rpt_family="L2"
repeat_region   complement(22969. .23401)
                /rpt_family="Tigger2a"
repeat_region   24050. .24410
                /rpt_family="MER57B"
repeat_region   25572. .25686
                /rpt_family="L1PB3"
repeat_region   25687. .26120
                /rpt_family="GA-rich"
repeat_region   26146. .26526
                /rpt_family="MER31-internal"
repeat_region   complement(26712. .27015)
                /rpt_family="AluY"

Alignment Scores:
Pred. No.:              1.89e+04        Length:         95546
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     8               Gaps:           0

US-09-858-332G-28 (1-6) x AC022816 (1-95546)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      32360 CACAACCATAACCATAAC 32377

RESULT 547
AX652128/c
LOCUS       AX652128               95769 bp    DNA     linear   PAT 22-MAR-2003
DEFINITION  Sequence 1979 from Patent WO03000898.
ACCESSION   AX652128
VERSION     AX652128.1  GI:29154942
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Chang,H.S., Chen,W., Cooper,B., Glazebrook,J., Goff,S.A., Hou,Y.M.,
            Katagiri,F., Quan,S., Tao,Y., Whitham,S., Xie,Z., Zhu,T. and Zou,G.
  TITLE     Plant genes involved in defense against pathogens
  JOURNAL   Patent: WO 03000898-A 1979 03-JAN-2003;
            Syngenta Participations AG (CH)
FEATURES             Location/Qualifiers
     source          1. .95769
                     /organism="Arabidopsis thaliana"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:3702"
ORIGIN

Alignment Scores:
Pred. No.:              1.89e+04        Length:         95769
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x AX652128 (1-95769)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      19488 CACAATCATAATCACAAT 19471

RESULT 548
AC013430/c
LOCUS       AC013430               95769 bp    DNA     linear   PLN 26-MAY-2000
DEFINITION  Genomic sequence for Arabidopsis thaliana BAC F3F9 from chromosome
            I, complete sequence.
ACCESSION   AC013430
```

```
VERSION     AC013430.5  GI:8096768
KEYWORDS    HTG.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 95769)
  AUTHORS   Chao,Q., Brooks,S., Buehler,E., Johnson-Hopson,C., Khan,S., Kim,C.,
            Shinn,P., Altafi,H., Bei,Q., Chin,C., Chiou,J., Choi,E., Conn,L.,
            Conway,A., Gonzales,A., Hansen,N., Howng,B., Koo,T., Lam,B.,
            Lee,J., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S., Mukharsky,N.,
            Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A.,
            Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G., Federspiel,N.A.,
            Theologis,A. and Ecker,J.R.
  TITLE     Genomic sequence for Arabidopsis thaliana BAC F3F9 from chromosome
            I
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 95769)
  AUTHORS   Ecker,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-1999) Arabidopsis thaliana Genome Center,
            Department of Biology, University of Pennsylvania, 38th Street and
            Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE   3  (bases 1 to 95769)
  AUTHORS   Ecker,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-2000) Arabidopsis thaliana Genome Center,
            Department of Biology, University of Pennsylvania, 38th Street and
            Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE   4  (bases 1 to 95769)
  AUTHORS   Cheuk,R., Shinn,P., Brooks,S., Buehler,E., Chao,Q.,
            Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bei,B., Chin,C.,
            Chiou,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N.,
            Howing,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J.,
            Liu,S., Mukharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H.,
            Schwartz,J., Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M.,
            Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAY-2000) Arabidopsis thaliana Genome Center,
            Department of Biology, University of Pennsylvania, 38th and
            Hamilton Walk, Philadelphia, PA 19104-6018, USA
REFERENCE   5  (bases 1 to 95769)
  AUTHORS   Cheuk,R., Shinn,P., Brooks,S., Buehler,E., Chao,Q.,
            Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bei,B., Chin,C.,
            Chiou,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N.,
            Howing,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J.,
            Liu,S., Mukharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H.,
            Schwartz,J., Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M.,
            Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAY-2000) Arabidopsis thaliana Genome Center,
            Department of Biology, University of Pennsylvania, 38th and
            Hamilton Walk, Philadelphia, PA 19104-6018, USA
COMMENT     On May 26, 2000 this sequence version replaced gi:6921155.
FEATURES             Location/Qualifiers
     source          1. .95769
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /chromosome="1"
                     /clone="F3F9"
     CDS             join(71. .402,574. .649,733. .933)
                     /note="similar to oxidosqualene cyclase dbj|BAA33462.1"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F3F9.1"
                     /protein_id="AAF71792.1"
                     /db_xref="GI:8052528"
                     /translation="MGHVQLLTPCFYSFPPFSFFLLFFLFNAKNVFILISVDLTVGTV
                     FELEAKVEQVIPPVKVEDGESTTHETRRMRYGEEFLSSRLCRPPMVTGLEKSLDLSSS
                     FLHWYVSSSFIGHLEEIFDAEHREEILRYIYRHLNDDGGWGLHVEGKSFMFCTALNYI
                     CLRILREGPDEGREYACKLTSKMPWIVTFLLIYNNESCFPSI"
     CDS             complement(join(3438. .3746,3868. .4032,4110. .4664))
                     /note="similar to F-box domain gb|AAF14684.1"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F3F9.2"
                     /protein_id="AAF71813.1"
                     /db_xref="GI:8052549"
                     /translation="MLASWIQSLTSMKHTKNLILANFNGRFAPTGEDNVLKLSTETFN
                     HPSLRTLSLYQYILQDAQAFNNCKNLKYLDLFQIFAEVDVFNEVISSCPSLEVLILQI
                     IFFNPSGHLKIDHKTLRILSMSCNQIDSIEVRAARLDILSIEYIPCESDNVVLEIPRL
                     QFGRNYWVAGRLLPHTSINISCPPQKQESNGMVKWDTNYATSPASLSVSLDLKNPREV
                     EVLRKILAVWTEKMIEVEISFKNNNPLGEEGNSSDGGAQNNLWEKAEPFPNADFRVDT
                     IWMHNFKGSNKEQFALASRFVMQKTVMKKMMIKTSFDEKKKKEIEAAVAKLKELPKGN
                     EELSIECF"
     CDS             join(7138. .7400,7491. .7887)
                     /note="unknown protein; similar to EST gb|AI999747.1"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F3F9.3"
                     /protein_id="AAF71793.1"
                     /db_xref="GI:8052529"
                     /translation="METGLNILKLSLCVSLVVVGSYAQAPAPWNPSNGFRPGTCDHYE
                     CPTYKLVEAGYGFEIRMYDAALWISTSPIPSLSMTQATKTGFRRLNRYIEGDNKSNVK
                     MNMTAPVIAQATPGRSVYTVSLYLPKKNQQNPPQADDLHVRSTKPTYVAVRQIGGYVS
                     NNVAKDEAAALMESLRDSNWILPIEKSKGKLPAYFLAVYNPPSHTTARVINEIMVPFN
                     M"
     CDS             join(8524. .8759,8846. .9287)
                     /note="hypothetical protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F3F9.4"
                     /protein_id="AAF71794.1"
                     /db_xref="GI:8052530"
                     /translation="MEAALSIFKLSFLLTLLSGGSDLIDSLPPNCNRAECPSYEVVHA
                     GNGYEIHRYNTTVWISTEPIQDISLNEASGNGWNQLSDYMNGNNDYHQRIEIALPYIT
                     QVSQNLSTFIVSFFVPKAFQPDPPPGNNLHVQRWDSRYVAVKQISGYVADHKIGKQVA
                     ELKASLQGTVWAKAIEKSRETGGVGSAWAYTVAQFSWPFQWSQRVNEIWFPPFEMEDEE
                     TVSIVTQ"
     CDS             join(13444. .13814,13956. .14310,14399. .14662)
                     /note="similar to dioxygenase protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F3F9.5"
                     /protein_id="AAF71795.1"
                     /db_xref="GI:8052531"
                     /translation="MAVLSKPVAIPKSGFSLIPVIDMSDPESKHALVKACEDFGFFKV
                     INHGVSAELVSVLEHETVDFFSLPKSEKTQVAGYPFGYGNSKIGRNGDVGWVEYLLMN
                     ANHDSGSGPLFPSLLKSPGTFRNALEEYTTSVRKMTFDVLEKITDGLGIKPRNTLSKL
                     VSDQNTDSILRLNHYPPCPLSNKKTNGGKNVIGFGEHTDPQIISVLRSNNTSGLQINL
                     NDGSWISVPPDHTSFFFNVGDSLQVMTNGRFKSVRHRVLANCKKSRVSMIYFAGPSLT
                     QRIAPLTCLIDNEDERLYEEFTWSEYKNSTYNSRLSDNRLQQFERKTIKNLLN"
     CDS             complement(join(15755. .16715,16810. .16829))
                     /note="hypothetical protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F3F9.6"
                     /protein_id="AAF71812.1"
                     /db_xref="GI:8052548"
                     /translation="MPKPRLGIRGSELPQRQSPRLRTSLLSTSSDPHHLSRPITDRSP
                     KLGLDRRSPRSGGPHTDPLSQKKLGSRISGLESQLGQAQEELRLLKQQLAKAEAAKKR
                     AQEELHRKKSKKPNTPAPERDDIPGDGHQETDVFEVLDEKAKESEKTKNDELASKEDQ
                     INVLKARLYDLEKERVSLSEENETLKDQLKKTDTEMSCAKAKEDEIASKVSQIGEELE
                     ESNETTAKLKKKLESVEEAKETLEAEMKKLKVQTEQWRKAADAAAAVLSGGVEMNGRF
                     SEQCGSMEKHFAGRFVGSPGMADDSDDGSGKRKSSGKKMFGDLWKKKGQK"
     CDS             complement(join(19162. .19689,19767. .19833,20044. .20164,
                     20234. .20378,20482. .20623,20693. .20738,20826. .20877,
                     20965. .21156))
                     /note="unknown protein; similar to ESTs gb|T41672.1,
                     gb|AI992710.1, and gb|AA586078.1"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F3F9.7"
                     /protein_id="AAF71811.1"
```

```
                     /db_xref="GI:8052547"
                     /translation="MGNKLGRKRQVVEERYTKPQGLYVNKDVDVKKLRKLIVESKLAP
                     CYPGDDESCHDLEECPICFLYYPSLNRSRCCMKSICTECFLQMKNPNSARPTQKFLIN
                     LMGFFPFVSSLTFMRCPFCKTPNYAVEYRGVKSKEEKGIEQVEEQRVIEAKIRMRQKE
                     MQDDEEKMQKRLESCSSSTSAMTGEMEYGSTSGFLFVPWFAAISYNSLMDDGEIAPSQ
                     NASVVRQHSRPRGNREDEVDVDLEELMVMEAIWLSVQETGTQRNSASGEITSSRQYVT
                     DNHSYVSSPPRVTPIVEPATPSSSSGGLSCAISALAERQMVGESSSHNHNHNVNVSSY
                     SMLPGNCDSYYDIEQEVDGIDNHHHHRHHYEMGETGSSNSYVSSYMTGEGFHNFPPPP
                     PLVIVPESFEEQMMMAMAVSMAEVHATTTCAPTEVTWQ"
     CDS             complement(join(22808. .23063,23118. .23244,23511. .23706))
                     /note="hypothetical protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F3F9.8"
                     /protein_id="AAF71810.1"
                     /db_xref="GI:8052546"
                     /translation="MSGRGKVKSEPMKVVFINTQYVETDARSFKTVVQELTGKNAIVA
                     AGPFDSPSAFDGRCYDGGSKIGPITNVVFWKVIMSYLSVRPMSPNFKRAYYSFHMLFK
                     RTMKPKTNYQNKKCECVTKALKATSRLRSQASTMGFSLWGVFATQVQLGITFFSIFFI
                     YLYFMFFFSCKIMFLTKRILISLKLVISLKTV"
     CDS             join(26752. .26907,27020. .27871,27923. .28153)
                     /note="similar to exopolygalacturonase precursor
                     sp|Q00293|PGLX_ASPTU"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F3F9.9"
                     /protein_id="AAF71796.1"
                     /db_xref="GI:8052532"
                     /translation="MVSTISGFSLILFIAAVASSISAAPSAALVGRKVFDVRSYGARG
                     DGKTDNTMAFTKAWKDACQWKGLPRVYIPFGTFYLGAVAFTGPCKSRISFIIKGTLLA
                     PKDPNAIKQDSWIIFRYVDYLTVSGGGILDGQGSYSWPLNNCRQTHNCRALPMNMGFQ
                     FVRPSRLTRIKSINSKMGHLNFFSVQHFDITRVNIKAPGDSPNTDGIKIGSSNHMKIH
                     HVDIATGDDCIAILSGTFNLDINKVNCGPGHGISVGSLGKFKGEKSVQGLIVRNSIFN
                     GTSNGVRIKTWPSPGEPNLVSNFLFKNLQMIDVQSPINIDQRYCPNPPCSFQVTSLTR
                     NKSFSKIQIRDVKFQNIWGTSTAKEAVKLQCSKNVPCKNVQLFNINIVHRGRDGPATS
                     VCENVGGWIGGKISPPSCIR"
     CDS             33566. .35539
                     /note="similar to 9-cis-epoxycarotenoid dioxygenase
                     gb|AAF26356.1"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F3F9.10"
                     /protein_id="AAF71797.1"
                     /db_xref="GI:8052533"
                     /translation="MTIITIISGMYIYSLLSQDAHHSQYGQNTNLVLKKPIPKPQTAA
                     FNQESTMASTTLLPSTSTQFLDRTFSTSSSSSRPKLQSLSFSSTLRNKKLVVPCYVSS
                     SVNKKSSVSSSLQSPTFKPPSWKKLCNDVTNLIPKTTNQNPKLNPVQRTAAMVLDAVE
                     NAMISHERRRHPHPKTADPAVQIAGNFFPVPEKPVVHNLPVTGTVPECIQGVYVRNGA
                     NPLHKPVSGHHLFDGDGMVHAVRFDNGSVSYACRFTETNRLVQERECGRPVFPKAIGE
                     LHGHLGIAKLMLFNTRGLFGLVDPTGGLGVANAGLVYFNGHLLAMSEDDLPYHVKVTQ

Alignment Scores:
Pred. No.:              1.89e+04        Length:         95769
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x AC013430 (1-95769)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      19488 CACAATCATAATCACAAT 19471

RESULT 549
AC149801
LOCUS       AC149801               96221 bp    DNA     linear   PLN 27-AUG-2004
DEFINITION  Medicago truncatula chromosome 2 BAC clone mth2-32i15, complete
            sequence.
ACCESSION   AC149801
VERSION     AC149801.3  GI:51571987
KEYWORDS    HTG.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 96221)
  AUTHORS   Town,C.D., Tallon,L.J., Arbogast,T., Althoff,R., Hine,E.,
            Monaghan,E., Smith,S.A., Utterback,T., Feldblyum,T., Koo,H. and
            Cheung,F.
  TITLE     Medicago truncatula BAC genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 96221)
  AUTHORS   Town,C.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JUN-2004) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 96221)
  AUTHORS   Town,C.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-AUG-2004) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   4  (bases 1 to 96221)
  AUTHORS   Town,C.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-2004) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
COMMENT     On Aug 27, 2004 this sequence version replaced gi:51556317.
            EXCEPTION:  Region 53267-53267 with a mononucleotide T repeat of
            9-11 nucleotides.
FEATURES             Location/Qualifiers
     source          1. .96221
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3880"
                     /chromosome="2"
                     /clone="mth2-32i15"
ORIGIN

Alignment Scores:
Pred. No.:              1.9e+04         Length:         96221
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x AC149801 (1-96221)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      42611 CACAACCATAACCATAAC 42628

RESULT 550
AP007568
LOCUS       AP007568               96315 bp    DNA     linear   HTG 28-DEC-2004
DEFINITION  Lotus corniculatus var. japonicus chromosome 3 clone LjT07O23, ***
            SEQUENCING IN PROGRESS ***, 30 unordered pieces.
ACCESSION   AP007568
VERSION     AP007568.1  GI:56805882
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1
  AUTHORS   Kaneko,T., Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Structural Analysis of a Lotus japonicus Genome. XI. Sequence
            Features and Mapping of Nine hundred twenty-one TAC Clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 96315)
```

AUTHORS Sato,S.
TITLE Direct Submission
JOURNAL Submitted (26-OCT-2004) Shusei Sato, Kazusa DNA Research Institute, Department of Plant Gene Research; 2-6-7 Kazusa-kamatari, Kisarazu, Chiba, 292-0818, Japan (E-mail:ssato@kazusa.or.jp, URL:http://www.kazusa.or.jp/, Tel:81-438-52-3935(ex.2337), Fax:81-438-52-3934)

COMMENT
* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*	1	7116: contig of 7116 bp in length
*	7117	7216: gap of unknown length
*	7217	8383: contig of 1167 bp in length
*	8384	8483: gap of unknown length
*	8484	10149: contig of 1666 bp in length
*	10150	10249: gap of unknown length
*	10250	11821: contig of 1572 bp in length
*	11822	11921: gap of unknown length
*	11922	13431: contig of 1510 bp in length
*	13432	13531: gap of unknown length
*	13532	14457: contig of 926 bp in length
*	14458	14557: gap of unknown length
*	14558	16184: contig of 1627 bp in length
*	16185	16284: gap of unknown length
*	16285	17525: contig of 1241 bp in length
*	17526	17625: gap of unknown length
*	17626	18667: contig of 1042 bp in length
*	18668	18767: gap of unknown length
*	18768	20138: contig of 1371 bp in length
*	20139	20238: gap of unknown length
*	20239	21315: contig of 1077 bp in length
*	21316	21415: gap of unknown length
*	21416	22603: contig of 1188 bp in length
*	22604	22703: gap of unknown length
*	22704	24132: contig of 1429 bp in length
*	24133	24232: gap of unknown length
*	24233	25443: contig of 1211 bp in length
*	25444	25543: gap of unknown length
*	25544	27103: contig of 1560 bp in length
*	27104	27203: gap of unknown length
*	27204	28858: contig of 1655 bp in length
*	28859	28958: gap of unknown length
*	28959	30236: contig of 1278 bp in length
*	30237	30336: gap of unknown length
*	30337	31767: contig of 1431 bp in length
*	31768	31867: gap of unknown length
*	31868	34193: contig of 2326 bp in length
*	34194	34293: gap of unknown length
*	34294	36695: contig of 2402 bp in length
*	36696	36795: gap of unknown length
*	36796	39856: contig of 3061 bp in length
*	39857	39956: gap of unknown length
*	39957	42889: contig of 2933 bp in length
*	42890	42989: gap of unknown length
*	42990	47078: contig of 4089 bp in length
*	47079	47178: gap of unknown length
*	47179	51924: contig of 4746 bp in length
*	51925	52024: gap of unknown length
*	52025	58254: contig of 6230 bp in length
*	58255	58354: gap of unknown length
*	58355	65462: contig of 7108 bp in length
*	65463	65562: gap of unknown length
*	65563	70424: contig of 4862 bp in length
*	70425	70524: gap of unknown length
*	70525	79059: contig of 8535 bp in length
*	79060	79159: gap of unknown length
*	79160	88451: contig of 9292 bp in length
*	88452	88551: gap of unknown length
*	88552	96315: contig of 7764 bp in length.

FEATURES Location/Qualifiers

source	1. .96315
	/organism="Lotus corniculatus var. japonicus"
	/mol_type="genomic DNA"
	/variety="japonicus"
	/db_xref="taxon:34305"
	/chromosome="3"
	/clone="LjT07O23"
	/clone_lib="LjT library"
	/note="TAC clone:TM0751, synonym:Lotus japonicus"
gap	7117. .7216
	/estimated_length=unknown
gap	8384. .8483
	/estimated_length=unknown
gap	10150. .10249
	/estimated_length=unknown
gap	11822. .11921
	/estimated_length=unknown
gap	13432. .13531
	/estimated_length=unknown
gap	14458. .14557
	/estimated_length=unknown
gap	16185. .16284
	/estimated_length=unknown
gap	17526. .17625
	/estimated_length=unknown
gap	18668. .18767
	/estimated_length=unknown
gap	20139. .20238
	/estimated_length=unknown
gap	21316. .21415
	/estimated_length=unknown
gap	22604. .22703
	/estimated_length=unknown
gap	24133. .24232
	/estimated_length=unknown
gap	25444. .25543
	/estimated_length=unknown
gap	27104. .27203
	/estimated_length=unknown
gap	28859. .28958
	/estimated_length=unknown
gap	30237. .30336
	/estimated_length=unknown
gap	31768. .31867
	/estimated_length=unknown
gap	34194. .34293
	/estimated_length=unknown
gap	36696. .36795
	/estimated_length=unknown
gap	39857. .39956
	/estimated_length=unknown
gap	42890. .42989
	/estimated_length=unknown
gap	47079. .47178
	/estimated_length=unknown
gap	51925. .52024
	/estimated_length=unknown
gap	58255. .58354
	/estimated_length=unknown
gap	65463. .65562
	/estimated_length=unknown
gap	70425. .70524
	/estimated_length=unknown
gap	79060. .79159
	/estimated_length=unknown
gap	88452. .88551
	/estimated_length=unknown

ORIGIN

Alignment Scores:
Pred. No.: 1.9e+04 Length: 96315

```
Score:                  42.00          Matches:       6
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     14             Gaps:          0

US-09-858-332G-28 (1-6) x AP007568 (1-96315)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      37895 CATAATCATAATCATAAT 37912

RESULT 551
AC136041
LOCUS       AC136041               96482 bp    DNA     linear   HTG 29-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-336D23, *** SEQUENCING IN PROGRESS
            ***, 49 unordered pieces.
ACCESSION   AC136041
VERSION     AC136041.1  GI:24417912
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 96482)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., Mcleod,M., Mcneill,T., Meenen,E., Milosavljevic,A.,
            Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K.,
            Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D.,
            Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G.,
            Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H.,
            Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A.,
            Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E.,
            Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y.,
            Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A.,
            Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S.,
            Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A.,
            Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E.,
            Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A.,
            Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S.,
            Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D.,
            Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J.,
            Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R.,
            Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S.,
            Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X.,
            Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R.,
            Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 96482)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: KDEN
                Center clone name: CH230-336D23
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 45003 bases at least Q40
                Consensus quality: 47218 bases at least Q30
                Consensus quality: 48140 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 49 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1080: contig of 1080 bp in length
            *     1081     1180: gap of unknown length
            *     1181     2467: contig of 1287 bp in length
            *     2468     2567: gap of unknown length
            *     2568     3592: contig of 1025 bp in length
            *     3593     3692: gap of unknown length
            *     3693     4710: contig of 1018 bp in length
            *     4711     4810: gap of unknown length
            *     4811     6025: contig of 1215 bp in length
            *     6026     6125: gap of unknown length
            *     6126     7252: contig of 1127 bp in length
            *     7253     7352: gap of unknown length
            *     7353     8527: contig of 1175 bp in length
            *     8528     8627: gap of unknown length
            *     8628     9982: contig of 1355 bp in length
            *     9983    10082: gap of unknown length
            *    10083    11446: contig of 1364 bp in length
            *    11447    11546: gap of unknown length
            *    11547    12734: contig of 1188 bp in length
            *    12735    12834: gap of unknown length
            *    12835    13881: contig of 1047 bp in length
            *    13882    13981: gap of unknown length
            *    13982    15178: contig of 1197 bp in length
            *    15179    15278: gap of unknown length
            *    15279    16646: contig of 1368 bp in length
            *    16647    16746: gap of unknown length
            *    16747    18078: contig of 1332 bp in length
            *    18079    18178: gap of unknown length
            *    18179    19996: contig of 1818 bp in length
            *    19997    20096: gap of unknown length
            *    20097    21636: contig of 1540 bp in length
            *    21637    21736: gap of unknown length
            *    21737    23417: contig of 1681 bp in length
            *    23418    23517: gap of unknown length
            *    23518    24899: contig of 1382 bp in length
            *    24900    24999: gap of unknown length
            *    25000    26166: contig of 1167 bp in length
            *    26167    26266: gap of unknown length
```

```
            *    26267    27823: contig of 1557 bp in length
            *    27824    27923: gap of unknown length
            *    27924    29256: contig of 1333 bp in length
            *    29257    29356: gap of unknown length
            *    29357    30666: contig of 1310 bp in length
            *    30667    30766: gap of unknown length
            *    30767    32052: contig of 1286 bp in length
            *    32053    32152: gap of unknown length
            *    32153    33814: contig of 1662 bp in length
            *    33815    33914: gap of unknown length
            *    33915    35605: contig of 1691 bp in length
            *    35606    35705: gap of unknown length
            *    35706    37529: contig of 1824 bp in length
            *    37530    37629: gap of unknown length
            *    37630    38893: contig of 1264 bp in length
            *    38894    38993: gap of unknown length
            *    38994    40451: contig of 1458 bp in length
            *    40452    40551: gap of unknown length
            *    40552    42100: contig of 1549 bp in length
            *    42101    42200: gap of unknown length
            *    42201    43550: contig of 1350 bp in length
            *    43551    43650: gap of unknown length
            *    43651    46314: contig of 2664 bp in length
            *    46315    46414: gap of unknown length
            *    46415    47501: contig of 1087 bp in length
            *    47502    47601: gap of unknown length
            *    47602    49348: contig of 1747 bp in length
            *    49349    49448: gap of unknown length
            *    49449    50726: contig of 1278 bp in length
            *    50727    50826: gap of unknown length
            *    50827    52732: contig of 1906 bp in length
            *    52733    52832: gap of unknown length
            *    52833    54227: contig of 1395 bp in length
            *    54228    54327: gap of unknown length
            *    54328    56879: contig of 2552 bp in length
            *    56880    56979: gap of unknown length
            *    56980    58390: contig of 1411 bp in length
            *    58391    58490: gap of unknown length
            *    58491    60859: contig of 2369 bp in length
            *    60860    60959: gap of unknown length
            *    60960    62822: contig of 1863 bp in length
            *    62823    62922: gap of unknown length
            *    62923    65534: contig of 2612 bp in length
            *    65535    65634: gap of unknown length
            *    65635    68112: contig of 2478 bp in length
            *    68113    68212: gap of unknown length
            *    68213    71933: contig of 3721 bp in length
            *    71934    72033: gap of unknown length
            *    72034    75676: contig of 3643 bp in length
            *    75677    75776: gap of unknown length
            *    75777    79583: contig of 3807 bp in length
            *    79584    79683: gap of unknown length
            *    79684    82875: contig of 3192 bp in length
            *    82876    82975: gap of unknown length
            *    82976    87254: contig of 4279 bp in length
            *    87255    87354: gap of unknown length
            *    87355    91674: contig of 4320 bp in length
            *    91675    91774: gap of unknown length
            *    91775    96482: contig of 4708 bp in length.
FEATURES             Location/Qualifiers
     source          1. .96482
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-336D23"
     gap             1081. .1180
                     /estimated_length=unknown
     gap             2468. .2567

Alignment Scores:
Pred. No.:              1.91e+04        Length:         96482
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AC136041 (1-96482)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      30202 CATAACCATAACCATAAC 30219

RESULT 552
AC013925
LOCUS       AC013925               97308 bp    DNA     linear   HTG 16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC013925
VERSION     AC013925.1  GI:6437410
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 97308)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10212147 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .97308
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
ORIGIN

Alignment Scores:
Pred. No.:              1.92e+04        Length:         97308
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AC013925 (1-97308)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      58948 CACAACCACAACCACAAC 58965

RESULT 553
AP008047
LOCUS       AP008047               97460 bp    DNA     linear   HTG 28-DEC-2004
DEFINITION  Lotus corniculatus var. japonicus chromosome 4 clone LjT46B12, ***
            SEQUENCING IN PROGRESS ***, 25 unordered pieces.
ACCESSION   AP008047
VERSION     AP008047.1  GI:56806355
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1
  AUTHORS   Kaneko,T., Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Structural Analysis of a Lotus japonicus Genome. XI. Sequence
            Features and Mapping of Nine hundred twenty-one TAC Clones
  JOURNAL   Unpublished
```

```
REFERENCE   2  (bases 1 to 97460)
  AUTHORS   Sato,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-2004) Shusei Sato, Kazusa DNA Research Institute,
            Department of Plant Gene Research; 2-6-7 Kazusa-kamatari, Kisarazu,
            Chiba, 292-0818, Japan (E-mail:ssato@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/, Tel:81-438-52-3935(ex.2337),
            Fax:81-438-52-3934)
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 25 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2945: contig of 2945 bp in length
            *     2946     3045: gap of unknown length
            *     3046     4014: contig of 969 bp in length
            *     4015     4114: gap of unknown length
            *     4115     5115: contig of 1001 bp in length
            *     5116     5215: gap of unknown length
            *     5216     5923: contig of 708 bp in length
            *     5924     6023: gap of unknown length
            *     6024     7550: contig of 1527 bp in length
            *     7551     7650: gap of unknown length
            *     7651     9278: contig of 1628 bp in length
            *     9279     9378: gap of unknown length
            *     9379    10687: contig of 1309 bp in length
            *    10688    10787: gap of unknown length
            *    10788    11965: contig of 1178 bp in length
            *    11966    12065: gap of unknown length
            *    12066    13430: contig of 1365 bp in length
            *    13431    13530: gap of unknown length
            *    13531    14632: contig of 1102 bp in length
            *    14633    14732: gap of unknown length
            *    14733    16551: contig of 1819 bp in length
            *    16552    16651: gap of unknown length
            *    16652    19724: contig of 3073 bp in length
            *    19725    19824: gap of unknown length
            *    19825    21775: contig of 1951 bp in length
            *    21776    21875: gap of unknown length
            *    21876    24012: contig of 2137 bp in length
            *    24013    24112: gap of unknown length
            *    24113    25664: contig of 1552 bp in length
            *    25665    25764: gap of unknown length
            *    25765    29356: contig of 3592 bp in length
            *    29357    29456: gap of unknown length
            *    29457    32846: contig of 3390 bp in length
            *    32847    32946: gap of unknown length
            *    32947    36901: contig of 3955 bp in length
            *    36902    37001: gap of unknown length
            *    37002    42640: contig of 5639 bp in length
            *    42641    42740: gap of unknown length
            *    42741    47502: contig of 4762 bp in length
            *    47503    47602: gap of unknown length
            *    47603    53395: contig of 5793 bp in length
            *    53396    53495: gap of unknown length
            *    53496    61363: contig of 7868 bp in length
            *    61364    61463: gap of unknown length
            *    61464    71244: contig of 9781 bp in length
            *    71245    71344: gap of unknown length
            *    71345    93230: contig of 21886 bp in length
            *    93231    93330: gap of unknown length
            *    93331    97460: contig of 4130 bp in length.
FEATURES             Location/Qualifiers
     source          1. .97460
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="genomic DNA"
                     /variety="japonicus"
                     /db_xref="taxon:34305"
                     /chromosome="4"
                     /clone="LjT46B12"
                     /clone_lib="LjT library"
                     /note="TAC clone:TM1399, synonym:Lotus japonicus"
     gap             2946. .3045
                     /estimated_length=unknown
     gap             4015. .4114
                     /estimated_length=unknown
     gap             5116. .5215
                     /estimated_length=unknown
     gap             5924. .6023
                     /estimated_length=unknown
     gap             7551. .7650
                     /estimated_length=unknown
     gap             9279. .9378
                     /estimated_length=unknown
     gap             10688. .10787
                     /estimated_length=unknown
     gap             11966. .12065
                     /estimated_length=unknown
     gap             13431. .13530
                     /estimated_length=unknown
     gap             14633. .14732
                     /estimated_length=unknown
     gap             16552. .16651
                     /estimated_length=unknown
     gap             19725. .19824
                     /estimated_length=unknown
     gap             21776. .21875
                     /estimated_length=unknown
     gap             24013. .24112
                     /estimated_length=unknown
     gap             25665. .25764
                     /estimated_length=unknown
     gap             29357. .29456
                     /estimated_length=unknown
     gap             32847. .32946
                     /estimated_length=unknown
     gap             36902. .37001
                     /estimated_length=unknown
     gap             42641. .42740
                     /estimated_length=unknown
     gap             47503. .47602
                     /estimated_length=unknown
     gap             53396. .53495
                     /estimated_length=unknown
     gap             61364. .61463
                     /estimated_length=unknown
     gap             71245. .71344
                     /estimated_length=unknown
     gap             93231. .93330
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              1.92e+04        Length:         97460
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AP008047 (1-97460)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      61678 CATAATCATAATCATAAC 61695

RESULT 554
AC024642/c
LOCUS       AC024642               97825 bp    DNA     linear   HTG 13-JUL-2000
DEFINITION  Homo sapiens chromosome 12 clone RP11-290M23 map 12, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC024642
VERSION     AC024642.2  GI:7329404
```

```
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 97825)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 12, clone RP11-290M23
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 97825)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 26, 2000 this sequence version replaced gi:7139855.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L6974
                Center clone name: 290_M_23
            --------------.
            * NOTE: This record contains 119 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      764: contig of 764 bp in length
            *      765      864: gap of 100 bp
            *      865     1592: contig of 728 bp in length
            *     1593     1692: gap of 100 bp
            *     1693     2404: contig of 712 bp in length
            *     2405     2504: gap of 100 bp
            *     2505     3215: contig of 711 bp in length
            *     3216     3315: gap of 100 bp
            *     3316     4053: contig of 738 bp in length
            *     4054     4153: gap of 100 bp
            *     4154     4898: contig of 745 bp in length
            *     4899     4998: gap of 100 bp
            *     4999     5739: contig of 741 bp in length
            *     5740     5839: gap of 100 bp
            *     5840     6582: contig of 743 bp in length
            *     6583     6682: gap of 100 bp
            *     6683     7410: contig of 728 bp in length
            *     7411     7510: gap of 100 bp
            *     7511     8236: contig of 726 bp in length
            *     8237     8336: gap of 100 bp
            *     8337     9072: contig of 736 bp in length
            *     9073     9172: gap of 100 bp
            *     9173     9910: contig of 738 bp in length
            *     9911    10010: gap of 100 bp
            *    10011    10723: contig of 713 bp in length
            *    10724    10823: gap of 100 bp
            *    10824    11538: contig of 715 bp in length
            *    11539    11638: gap of 100 bp
            *    11639    12353: contig of 715 bp in length
            *    12354    12453: gap of 100 bp
            *    12454    13191: contig of 738 bp in length
            *    13192    13291: gap of 100 bp
            *    13292    13997: contig of 706 bp in length
            *    13998    14097: gap of 100 bp
            *    14098    14845: contig of 748 bp in length
            *    14846    14945: gap of 100 bp
            *    14946    15675: contig of 730 bp in length
            *    15676    15775: gap of 100 bp
            *    15776    16508: contig of 733 bp in length
            *    16509    16608: gap of 100 bp
            *    16609    17338: contig of 730 bp in length
            *    17339    17438: gap of 100 bp
            *    17439    18156: contig of 718 bp in length
            *    18157    18256: gap of 100 bp
            *    18257    19002: contig of 746 bp in length
            *    19003    19102: gap of 100 bp
            *    19103    19835: contig of 733 bp in length
            *    19836    19935: gap of 100 bp
            *    19936    20652: contig of 717 bp in length
            *    20653    20752: gap of 100 bp
            *    20753    21504: contig of 752 bp in length
            *    21505    21604: gap of 100 bp
            *    21605    22338: contig of 734 bp in length
            *    22339    22438: gap of 100 bp
            *    22439    23193: contig of 755 bp in length
            *    23194    23293: gap of 100 bp
            *    23294    24043: contig of 750 bp in length
            *    24044    24143: gap of 100 bp
            *    24144    24887: contig of 744 bp in length
            *    24888    24987: gap of 100 bp
            *    24988    25715: contig of 728 bp in length
            *    25716    25815: gap of 100 bp
            *    25816    26556: contig of 741 bp in length
            *    26557    26656: gap of 100 bp
            *    26657    27403: contig of 747 bp in length
            *    27404    27503: gap of 100 bp
            *    27504    28242: contig of 739 bp in length
            *    28243    28342: gap of 100 bp
            *    28343    29073: contig of 731 bp in length
            *    29074    29173: gap of 100 bp
            *    29174    29896: contig of 723 bp in length
            *    29897    29996: gap of 100 bp
            *    29997    30707: contig of 711 bp in length
            *    30708    30807: gap of 100 bp
            *    30808    31525: contig of 718 bp in length
            *    31526    31625: gap of 100 bp
            *    31626    32341: contig of 716 bp in length
            *    32342    32441: gap of 100 bp
            *    32442    33179: contig of 738 bp in length
            *    33180    33279: gap of 100 bp
            *    33280    34024: contig of 745 bp in length
            *    34025    34124: gap of 100 bp
            *    34125    34837: contig of 713 bp in length
            *    34838    34937: gap of 100 bp
            *    34938    35680: contig of 743 bp in length
            *    35681    35780: gap of 100 bp
            *    35781    36507: contig of 727 bp in length
```

```
            *    36508    36607: gap of 100 bp
            *    36608    37347: contig of 740 bp in length
            *    37348    37447: gap of 100 bp
            *    37448    38171: contig of 724 bp in length
            *    38172    38271: gap of 100 bp
            *    38272    38992: contig of 721 bp in length
            *    38993    39092: gap of 100 bp
            *    39093    39806: contig of 714 bp in length
            *    39807    39906: gap of 100 bp
            *    39907    40623: contig of 717 bp in length
            *    40624    40723: gap of 100 bp
            *    40724    41457: contig of 734 bp in length
            *    41458    41557: gap of 100 bp
            *    41558    42304: contig of 747 bp in length
            *    42305    42404: gap of 100 bp
            *    42405    43161: contig of 757 bp in length
            *    43162    43261: gap of 100 bp
            *    43262    44009: contig of 748 bp in length
            *    44010    44109: gap of 100 bp
            *    44110    44852: contig of 743 bp in length
            *    44853    44952: gap of 100 bp
            *    44953    45666: contig of 714 bp in length
            *    45667    45766: gap of 100 bp
            *    45767    46496: contig of 730 bp in length
            *    46497    46596: gap of 100 bp
            *    46597    47337: contig of 741 bp in length
            *    47338    47437: gap of 100 bp
            *    47438    48184: contig of 747 bp in length
            *    48185    48284: gap of 100 bp
            *    48285    49017: contig of 733 bp in length
            *    49018    49117: gap of 100 bp
            *    49118    49846: contig of 729 bp in length
            *    49847    49946: gap of 100 bp
            *    49947    50659: contig of 713 bp in length
            *    50660    50759: gap of 100 bp
            *    50760    51501: contig of 742 bp in length
            *    51502    51601: gap of 100 bp
            *    51602    52351: contig of 750 bp in length
            *    52352    52451: gap of 100 bp
            *    52452    53164: contig of 713 bp in length
            *    53165    53264: gap of 100 bp
            *    53265    53984: contig of 720 bp in length
            *    53985    54084: gap of 100 bp
            *    54085    54823: contig of 739 bp in length
            *    54824    54923: gap of 100 bp
            *    54924    55658: contig of 735 bp in length
            *    55659    55758: gap of 100 bp
            *    55759    56500: contig of 742 bp in length
            *    56501    56600: gap of 100 bp

Alignment Scores:
Pred. No.:                1.93e+04        Length:          97825
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       14              Gaps:            0

US-09-858-332G-28 (1-6) x AC024642 (1-97825)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        356 CACAACCACAACCACAAC 339

RESULT 555
AY438979/c
LOCUS       AY438979               98039 bp    DNA     linear   PRI 29-OCT-2003
DEFINITION  Homo sapiens transforming growth factor, beta 2 (TGFB2) gene,
            complete cds.
ACCESSION   AY438979
VERSION     AY438979.1  GI:37953286
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 98039)
  AUTHORS   Rieder,M.J., Livingston,R.J., Daniels,M.R., Chung,M.-W.,
            Miyamoto,K.E., Nguyen,C.P., Nguyen,D.A., Poel,C.L., Robertson,P.D.,
            Schackwitz,W.S., Sherwood,J.K., Witrak,L.A. and Nickerson,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-OCT-2003) Genome Sciences, University of Washington,
            1705 NE Pacific, Seattle, WA 98195, USA
COMMENT     To cite this work please use:  NIEHS-SNPs, Environmental Genome
            Project, NIEHS ES15478, Department of Genome Sciences, Seattle, WA
            (URL: http://egp.gs.washington.edu).
FEATURES             Location/Qualifiers
     source          1. .98039
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     variation       486. .489
                     /frequency="0.12"
                     /replace=""
     variation       572
                     /frequency="0.17"
                     /replace="c"
     variation       799
                     /frequency="0.03"
                     /replace="g"
     misc_feature    884. .1189
                     /note="Region not scanned for variation"
     gene            1925. .97039
                     /gene="TGFB2"
     mRNA            join(1925. .2450,60574. .60737,89491. .89623,89747. .89857,
                     91379. .91556,92752. .92905,96615. .97039)
                     /gene="TGFB2"
                     /product="transforming growth factor, beta 2"
     variation       1992. .1995
                     /gene="TGFB2"
                     /frequency="0.56"
                     /replace=""
     CDS             join(2105. .2450,60574. .60737,89491. .89623,89747. .89857,
                     91379. .91556,92752. .92905,96615. .96773)
                     /gene="TGFB2"
                     /codon_start=1
                     /product="transforming growth factor, beta 2"
                     /protein_id="AAR05442.1"
                     /db_xref="GI:37953287"
                     /translation="MHYCVLSAFLILHLVTVALSLSTCSTLDMDQFMRKRIEAIRGQI
                     LSKLKLTSPPEDYPEPEEVPPEVISIYNSTRDLLQEKASRRAAACERERSDEEYYAKE
                     VYKIDMPPFFPSENAIPPTFYRPYFRIVRFDVSAMEKNASNLVKAEFRVFRLQNPKAR
                     VPEQRIELYQILKSKDLTSPTQRYIDSKVVKTRAEGEWLSFDVTDAVHEWLHHKDRNL
                     GFKISLHCPCCTFVPSNNYIIPNKSEELEARFAGIDGTSTYTSGDQKTIKSTRKKNSG
                     KTPHLLLMLLPSYRLESQQTNRRKKRALDAAYCFRNVQDNCCLRPLYIDFKRDLGWKW
                     IHEPKGYNANFCAGACPYLWSSDTQHSRVLSLYNTINPEASASPCCVSQDLEPLTILY
                     YIGKTPKIEQLSNMIVKSCKCS"
     variation       2210
                     /gene="TGFB2"
                     /frequency="0.01"
                     /replace="c"
     variation       2376
                     /gene="TGFB2"
                     /frequency="0.01"
                     /replace="a"
     variation       2648
                     /gene="TGFB2"
                     /frequency="0.03"
                     /replace="g"
     variation       2665
                     /gene="TGFB2"
                     /frequency="0.01"
                     /replace="g"
     variation       2775
                     /gene="TGFB2"
```

```
                /frequency="0.01"
                /replace="g"
variation       2904
                /gene="TGFB2"
                /frequency="0.16"
                /replace="a"
variation       3056
                /gene="TGFB2"
                /frequency="0.43"
                /replace="g"
variation       3072
                /gene="TGFB2"
                /frequency="0.01"
                /replace="c"
variation       3670
                /gene="TGFB2"
                /frequency="0.16"
                /replace="g"
variation       3829
                /gene="TGFB2"
                /frequency="0.01"
                /replace="g"
variation       4342
                /gene="TGFB2"
                /frequency="0.04"
                /replace="g"
variation       4529
                /gene="TGFB2"
                /frequency="0.01"
                /replace="t"
variation       4623
                /gene="TGFB2"
                /frequency="0.01"
                /replace="c"
misc_feature    4716. .16780
                /gene="TGFB2"
                /note="Region not scanned for variation"
repeat_region   5728. .5916
                /rpt_family="MER1_type"
                /rpt_type=dispersed
repeat_region   8278. .8407
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   8434. .8561
                /rpt_family="Tc2"
                /rpt_type=dispersed
repeat_region   8829. .9018
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   9019. .9316
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   9317. .9403
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   9864. .10001
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   10305. .10602
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   10952. .11119
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   11142. .11334
                /rpt_family="Tc2"
                /rpt_type=dispersed
repeat_region   11628. .11921
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   12333. .12641
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   13952. .14243
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   15455. .15647
                /rpt_family="MER1_type"
                /rpt_type=dispersed
repeat_region   15872. .16099
                /rpt_family="L1"
                /rpt_type=dispersed
repeat_region   16259. .16555
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       17051
                /gene="TGFB2"
                /frequency="0.02"
                /replace="c"
variation       17220
                /gene="TGFB2"
                /frequency="0.01"
                /replace="c"
variation       17326
                /gene="TGFB2"
                /frequency="0.02"
                /replace="a"
variation       17670
                /gene="TGFB2"
                /frequency="0.01"
                /replace="g"
variation       17716
                /gene="TGFB2"
                /frequency="0.01"
                /replace="g"
variation       17740
                /gene="TGFB2"
                /frequency="0.02"
                /replace="t"
variation       17879
                /gene="TGFB2"
                /frequency="0.02"
                /replace="a"
variation       17976
                /gene="TGFB2"
                /frequency="0.01"
                /replace="g"
variation       17979
                /gene="TGFB2"
                /frequency="0.01"

Alignment Scores:
Pred. No.:              1.94e+04        Length:         98039
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     8               Gaps:           0

US-09-858-332G-28 (1-6) x AY438979 (1-98039)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      95637 CATAATCATAATCACAAT 95620

RESULT 556
AP008210_354
WPCOMMENT
  Sequence split into 355 fragments  LOCUS AP008210 Accession AP008210
     Fragment Name      Begin     End
     AP008210_000           1     110000
     AP008210_001      100001     210000
     AP008210_002      200001     310000
     AP008210_003      300001     410000
     AP008210_004      400001     510000
     AP008210_005      500001     610000
```

AP008210_006	600001	710000
AP008210_007	700001	810000
AP008210_008	800001	910000
AP008210_009	900001	1010000
AP008210_010	1000001	1110000
AP008210_011	1100001	1210000
AP008210_012	1200001	1310000
AP008210_013	1300001	1410000
AP008210_014	1400001	1510000
AP008210_015	1500001	1610000
AP008210_016	1600001	1710000
AP008210_017	1700001	1810000
AP008210_018	1800001	1910000
AP008210_019	1900001	2010000
AP008210_020	2000001	2110000
AP008210_021	2100001	2210000
AP008210_022	2200001	2310000
AP008210_023	2300001	2410000
AP008210_024	2400001	2510000
AP008210_025	2500001	2610000
AP008210_026	2600001	2710000
AP008210_027	2700001	2810000
AP008210_028	2800001	2910000
AP008210_029	2900001	3010000
AP008210_030	3000001	3110000
AP008210_031	3100001	3210000
AP008210_032	3200001	3310000
AP008210_033	3300001	3410000
AP008210_034	3400001	3510000
AP008210_035	3500001	3610000
AP008210_036	3600001	3710000
AP008210_037	3700001	3810000
AP008210_038	3800001	3910000
AP008210_039	3900001	4010000
AP008210_040	4000001	4110000
AP008210_041	4100001	4210000
AP008210_042	4200001	4310000
AP008210_043	4300001	4410000
AP008210_044	4400001	4510000
AP008210_045	4500001	4610000
AP008210_046	4600001	4710000
AP008210_047	4700001	4810000
AP008210_048	4800001	4910000
AP008210_049	4900001	5010000
AP008210_050	5000001	5110000
AP008210_051	5100001	5210000
AP008210_052	5200001	5310000
AP008210_053	5300001	5410000
AP008210_054	5400001	5510000
AP008210_055	5500001	5610000
AP008210_056	5600001	5710000
AP008210_057	5700001	5810000
AP008210_058	5800001	5910000
AP008210_059	5900001	6010000
AP008210_060	6000001	6110000
AP008210_061	6100001	6210000
AP008210_062	6200001	6310000
AP008210_063	6300001	6410000
AP008210_064	6400001	6510000
AP008210_065	6500001	6610000
AP008210_066	6600001	6710000
AP008210_067	6700001	6810000
AP008210_068	6800001	6910000
AP008210_069	6900001	7010000
AP008210_070	7000001	7110000
AP008210_071	7100001	7210000
AP008210_072	7200001	7310000
AP008210_073	7300001	7410000
AP008210_074	7400001	7510000
AP008210_075	7500001	7610000
AP008210_076	7600001	7710000
AP008210_077	7700001	7810000
AP008210_078	7800001	7910000
AP008210_079	7900001	8010000
AP008210_080	8000001	8110000
AP008210_081	8100001	8210000
AP008210_082	8200001	8310000
AP008210_083	8300001	8410000
AP008210_084	8400001	8510000
AP008210_085	8500001	8610000
AP008210_086	8600001	8710000
AP008210_087	8700001	8810000
AP008210_088	8800001	8910000
AP008210_089	8900001	9010000
AP008210_090	9000001	9110000
AP008210_091	9100001	9210000
AP008210_092	9200001	9310000
AP008210_093	9300001	9410000
AP008210_094	9400001	9510000
AP008210_095	9500001	9610000
AP008210_096	9600001	9710000
AP008210_097	9700001	9810000
AP008210_098	9800001	9910000
AP008210_099	9900001	10010000
AP008210_100	10000001	10110000
AP008210_101	10100001	10210000
AP008210_102	10200001	10310000
AP008210_103	10300001	10410000
AP008210_104	10400001	10510000
AP008210_105	10500001	10610000
AP008210_106	10600001	10710000
AP008210_107	10700001	10810000
AP008210_108	10800001	10910000
AP008210_109	10900001	11010000
AP008210_110	11000001	11110000
AP008210_111	11100001	11210000
AP008210_112	11200001	11310000
AP008210_113	11300001	11410000
AP008210_114	11400001	11510000
AP008210_115	11500001	11610000
AP008210_116	11600001	11710000
AP008210_117	11700001	11810000
AP008210_118	11800001	11910000
AP008210_119	11900001	12010000
AP008210_120	12000001	12110000
AP008210_121	12100001	12210000
AP008210_122	12200001	12310000
AP008210_123	12300001	12410000
AP008210_124	12400001	12510000
AP008210_125	12500001	12610000
AP008210_126	12600001	12710000
AP008210_127	12700001	12810000
AP008210_128	12800001	12910000
AP008210_129	12900001	13010000
AP008210_130	13000001	13110000
AP008210_131	13100001	13210000
AP008210_132	13200001	13310000
AP008210_133	13300001	13410000
AP008210_134	13400001	13510000
AP008210_135	13500001	13610000
AP008210_136	13600001	13710000
AP008210_137	13700001	13810000
AP008210_138	13800001	13910000
AP008210_139	13900001	14010000
AP008210_140	14000001	14110000
AP008210_141	14100001	14210000
AP008210_142	14200001	14310000
AP008210_143	14300001	14410000
AP008210_144	14400001	14510000
AP008210_145	14500001	14610000
AP008210_146	14600001	14710000
AP008210_147	14700001	14810000
AP008210_148	14800001	14910000
AP008210_149	14900001	15010000
AP008210_150	15000001	15110000
AP008210_151	15100001	15210000

AP008210_152	15200001	15310000
AP008210_153	15300001	15410000
AP008210_154	15400001	15510000
AP008210_155	15500001	15610000
AP008210_156	15600001	15710000
AP008210_157	15700001	15810000
AP008210_158	15800001	15910000
AP008210_159	15900001	16010000
AP008210_160	16000001	16110000
AP008210_161	16100001	16210000
AP008210_162	16200001	16310000
AP008210_163	16300001	16410000
AP008210_164	16400001	16510000
AP008210_165	16500001	16610000
AP008210_166	16600001	16710000
AP008210_167	16700001	16810000
AP008210_168	16800001	16910000
AP008210_169	16900001	17010000
AP008210_170	17000001	17110000
AP008210_171	17100001	17210000
AP008210_172	17200001	17310000
AP008210_173	17300001	17410000
AP008210_174	17400001	17510000
AP008210_175	17500001	17610000
AP008210_176	17600001	17710000
AP008210_177	17700001	17810000
AP008210_178	17800001	17910000
AP008210_179	17900001	18010000
AP008210_180	18000001	18110000
AP008210_181	18100001	18210000
AP008210_182	18200001	18310000
AP008210_183	18300001	18410000
AP008210_184	18400001	18510000
AP008210_185	18500001	18610000
AP008210_186	18600001	18710000
AP008210_187	18700001	18810000
AP008210_188	18800001	18910000
AP008210_189	18900001	19010000
AP008210_190	19000001	19110000
AP008210_191	19100001	19210000
AP008210_192	19200001	19310000
AP008210_193	19300001	19410000
AP008210_194	19400001	19510000
AP008210_195	19500001	19610000
AP008210_196	19600001	19710000

Alignment Scores:

Pred. No.:	1.94e+04	Length:	98469
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	15	Gaps:	0

US-09-858-332G-28 (1-6) x AP008210_354 (1-98469)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      70702 CACAACCACAACCACAAC 70719
```

RESULT 557
AC133749/c
LOCUS AC133749 98760 bp DNA linear PRI 28-FEB-2003
DEFINITION Homo sapiens 12 BAC RP11-249J13 (Roswell Park Cancer Institute Human BAC Library) complete sequence.
ACCESSION AC133749
VERSION AC133749.2 GI:23346647
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 98760)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Merscher,S., Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M., Neal,D., Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kucherlapati,R., Weinstock,G. and Gibbs,R.
TITLE Direct Submission
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 98760)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (18-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 98760)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (30-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 4 (bases 1 to 98760)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (28-FEB-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Sep 30, 2002 this sequence version replaced gi:23096711.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:
http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

```
FEATURES             Location/Qualifiers
     source          1. .98760
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-249J13"
     misc_feature    1. .2004
                     /note="overlaps bases 92632. .94635 of clone AC139931"
                     /function="clone overlap"
     repeat_region   170. .452
                     /rpt_family="AluJo"
     repeat_region   1303. .1335
                     /rpt_family="AT_rich"
     repeat_region   1421. .1543
                     /rpt_family="L2"
     repeat_region   1841. .2154
                     /rpt_family="AluSx"
     repeat_region   3030. .3210
                     /rpt_family="LTR33"
     repeat_region   3831. .4038
                     /rpt_family="L2"
     repeat_region   complement(4039. .4342)
                     /rpt_family="AluJo"
     repeat_region   4343. .4401
                     /rpt_family="L2"
     repeat_region   4409. .4552
                     /rpt_family="(TA)n"
     repeat_region   complement(4556. .4591)
                     /rpt_family="L2"
     repeat_region   4617. .4649
                     /rpt_family="L2"
     misc_feature    4643. .5066
                     /note="size confirmed by restriction digest
                     424 bp"
                     /function="unresolved tandem repeat"
     repeat_region   4650. .4667
                     /rpt_family="(TA)n"
     repeat_region   4668. .5026
                     /rpt_family="(TA)n"
     repeat_region   complement(5060. .5132)
                     /rpt_family="L2"
     repeat_region   5133. .5452
                     /rpt_family="AluJo"
     repeat_region   complement(5453. .5576)
                     /rpt_family="L2"
     repeat_region   complement(5602. .5635)
                     /rpt_family="L2"
     repeat_region   complement(5649. .5845)
                     /rpt_family="AluSx"
     repeat_region   6505. .6803
                     /rpt_family="AluSc"
     repeat_region   6874. .7092
                     /rpt_family="MIR"
     repeat_region   complement(7633. .7817)
                     /rpt_family="MER3"
     repeat_region   8497. .8798
                     /rpt_family="AluSq"
     repeat_region   complement(9974. .10019)
                     /rpt_family="L1MB2"
     repeat_region   complement(10020. .10318)
                     /rpt_family="AluSx"
     repeat_region   complement(10319. .10375)
                     /rpt_family="L1MB2"
     repeat_region   10785. .10827
                     /rpt_family="AT_rich"
     repeat_region   complement(11710. .12023)
                     /rpt_family="AluSx"
     repeat_region   12322. .12544
                     /rpt_family="AluSq"
     repeat_region   complement(12619. .12967)
                     /rpt_family="MLT1J2"
     repeat_region   13613. .13778
                     /rpt_family="MIR"
     repeat_region   14331. .14438
                     /rpt_family="AT_rich"
     repeat_region   14455. .14474
                     /rpt_family="(TA)n"
     repeat_region   14475. .14775
                     /rpt_family="AluSx"
     repeat_region   14776. .14785
                     /rpt_family="(TA)n"
     repeat_region   14808. .14836
                     /rpt_family="AT_rich"
     repeat_region   15665. .15789
                     /rpt_family="Charlie3"
     repeat_region   complement(16435. .16720)
                     /rpt_family="AluJo"

Alignment Scores:
Pred. No.:              1.95e+04        Length:         98760
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     8               Gaps:           0

US-09-858-332G-28 (1-6) x AC133749 (1-98760)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      44125 CATAATCATAATCATAAT 44108

RESULT 558
TANN3_18
WPCOMMENT
  Sequence split into 19 fragments  LOCUS TANN3 Accession CR940352
     Fragment Name      Begin     End
     TANN3_00                1    110000
     TANN3_01           100001    210000
     TANN3_02           200001    310000
     TANN3_03           300001    410000
     TANN3_04           400001    510000
     TANN3_05           500001    610000
     TANN3_06           600001    710000
     TANN3_07           700001    810000
     TANN3_08           800001    910000
     TANN3_09           900001   1010000
```

```
     TANN3_10          1000001   1110000
     TANN3_11          1100001   1210000
     TANN3_12          1200001   1310000
     TANN3_13          1300001   1410000
     TANN3_14          1400001   1510000
     TANN3_15          1500001   1610000
     TANN3_16          1600001   1710000
     TANN3_17          1700001   1810000
     TANN3_18          1800001   1898942
Continuation (19 of 19) of TANN3 from base 1800001 (CR940352 Theileria annulata strain A

Alignment Scores:
Pred. No.:              1.95e+04       Length:         98942
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     14             Gaps:           0

US-09-858-332G-28 (1-6) x TANN3_18 (1-98942)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      83030 CACAACCACAACCACAAC 83047

RESULT 559
AP007508
LOCUS       AP007508               99104 bp    DNA     linear   HTG 28-DEC-2004
DEFINITION  Lotus corniculatus var. japonicus chromosome 1 clone LjT37O24, ***
            SEQUENCING IN PROGRESS ***, 21 unordered pieces.
ACCESSION   AP007508
VERSION     AP007508.1  GI:56805822
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1
  AUTHORS   Kaneko,T., Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Structural Analysis of a Lotus japonicus Genome. XI. Sequence
            Features and Mapping of Nine hundred twenty-one TAC Clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 99104)
  AUTHORS   Sato,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-2004) Shusei Sato, Kazusa DNA Research Institute,
            Department of Plant Gene Research; 2-6-7 Kazusa-kamatari, Kisarazu,
            Chiba, 292-0818, Japan (E-mail:ssato@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/, Tel:81-438-52-3935(ex.2337),
            Fax:81-438-52-3934)
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 21 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     5634: contig of 5634 bp in length
            *     5635     5734: gap of unknown length
            *     5735     6689: contig of 955 bp in length
            *     6690     6789: gap of unknown length
            *     6790     7816: contig of 1027 bp in length
            *     7817     7916: gap of unknown length
            *     7917    10345: contig of 2429 bp in length
            *    10346    10445: gap of unknown length
            *    10446    12278: contig of 1833 bp in length
            *    12279    12378: gap of unknown length
            *    12379    14486: contig of 2108 bp in length
            *    14487    14586: gap of unknown length
            *    14587    17300: contig of 2714 bp in length
            *    17301    17400: gap of unknown length
            *    17401    19679: contig of 2279 bp in length
            *    19680    19779: gap of unknown length
            *    19780    22506: contig of 2727 bp in length
            *    22507    22606: gap of unknown length
            *    22607    26758: contig of 4152 bp in length
            *    26759    26858: gap of unknown length
            *    26859    31414: contig of 4556 bp in length
            *    31415    31514: gap of unknown length
            *    31515    36792: contig of 5278 bp in length
            *    36793    36892: gap of unknown length
            *    36893    43034: contig of 6142 bp in length
            *    43035    43134: gap of unknown length
            *    43135    48654: contig of 5520 bp in length
            *    48655    48754: gap of unknown length
            *    48755    54036: contig of 5282 bp in length
            *    54037    54136: gap of unknown length
            *    54137    60104: contig of 5968 bp in length
            *    60105    60204: gap of unknown length
            *    60205    68811: contig of 8607 bp in length
            *    68812    68911: gap of unknown length
            *    68912    75850: contig of 6939 bp in length
            *    75851    75950: gap of unknown length
            *    75951    84281: contig of 8331 bp in length
            *    84282    84381: gap of unknown length
            *    84382    96464: contig of 12083 bp in length
            *    96465    96564: gap of unknown length
            *    96565    99104: contig of 2540 bp in length.
FEATURES             Location/Qualifiers
     source          1. .99104
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="genomic DNA"
                     /variety="japonicus"
                     /db_xref="taxon:34305"
                     /chromosome="1"
                     /clone="LjT37O24"
                     /clone_lib="LjT library"
                     /note="TAC clone:TM0676, synonym:Lotus japonicus"
     gap             5635. .5734
                     /estimated_length=unknown
     gap             6690. .6789
                     /estimated_length=unknown
     gap             7817. .7916
                     /estimated_length=unknown
     gap             10346. .10445
                     /estimated_length=unknown
     gap             12279. .12378
                     /estimated_length=unknown
     gap             14487. .14586
                     /estimated_length=unknown
     gap             17301. .17400
                     /estimated_length=unknown
     gap             19680. .19779
                     /estimated_length=unknown
     gap             22507. .22606
                     /estimated_length=unknown
     gap             26759. .26858
                     /estimated_length=unknown
     gap             31415. .31514
                     /estimated_length=unknown
     gap             36793. .36892
                     /estimated_length=unknown
     gap             43035. .43134
                     /estimated_length=unknown
     gap             48655. .48754
                     /estimated_length=unknown
     gap             54037. .54136
                     /estimated_length=unknown
     gap             60105. .60204
                     /estimated_length=unknown
     gap             68812. .68911
                     /estimated_length=unknown
     gap             75851. .75950
```

```
                /estimated_length=unknown
     gap        84282. .84381
                /estimated_length=unknown
     gap        96465. .96564
                /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              1.96e+04        Length:         99104
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AP007508 (1-99104)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      77046 CACAATCACAATCACAAT 77063

RESULT 560
AC026115/c
LOCUS       AC026115               99545 bp    DNA     linear   PRI 31-JUL-2001
DEFINITION  Homo sapiens 12 BAC RP11-1026F22 (Roswell Park Cancer Institute
            Human BAC Library) complete sequence.
ACCESSION   AC026115
VERSION     AC026115.27  GI:15042746
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 99545)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D.,
            Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D.,
            Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J.,
            Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W.,
            Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,
            Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O.,
            Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F.,
            Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R.,
            Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P.,
            Meador,M., Mei,G., Merscher,S., Metzker,M., Miller,A., Miner,G.,
            Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M.,
            Morris,S., Moser,M., Neal,D., Nelson,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C.,
            Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A.,
            Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A.,
            Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B.,
            Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D.,
            Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R.,
            Washington,C., Watlington,S., Williams,G., Williamson,A.,
            Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J.,
            Zorrilla,S., Kucherlapati,R., Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 99545)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAR-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 99545)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JUL-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 31, 2001 this sequence version replaced gi:15022624.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
               STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
               Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
               Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
            Reports of lowest quality individual bases and measures of base
            quality are listed below. Description of the metrics can be found
            at URL:
            http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

            QUALSTAT-REPORT.
FEATURES            Location/Qualifiers
     source         1. .99545
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /chromosome="12"
                    /clone="RP11-1026F22"
     repeat_region  complement(473. .673)
                    /rpt_family="MIR"
     repeat_region  694. .1535
                    /rpt_family="L2"
     STS            1335. .1586
                    /standard_name="163962"
     repeat_region  complement(1627. .1940)
                    /rpt_family="AluY"
     repeat_region  complement(2112. .2436)
                    /rpt_family="L2"
```

```
repeat_region   complement(2504. .2997)
                /rpt_family="L2"
repeat_region   2998. .3314
                /rpt_family="AluJo"
repeat_region   complement(3315. .3503)
                /rpt_family="L2"
repeat_region   3683. .3703
                /rpt_family="AT_rich"
repeat_region   3855. .4132
                /rpt_family="AluSq"
repeat_region   4151. .4212
                /rpt_family="MIR"
repeat_region   complement(4698. .5012)
                /rpt_family="AluJo"
repeat_region   complement(5178. .5464)
                /rpt_family="AluJb"
repeat_region   complement(6499. .6808)
                /rpt_family="AluSx"
repeat_region   6826. .7179
                /rpt_family="L1"
STS             7093. .7268
                /standard_name="83999"
repeat_region   7180. .9210
                /rpt_family="L1MB2"
repeat_region   complement(9211. .9545)
                /rpt_family="THE1B"
repeat_region   complement(10615. .10919)
                /rpt_family="AluSx"
repeat_region   11059. .11373
                /rpt_family="AluY"
repeat_region   complement(11740. .12083)
                /rpt_family="THE1B"
repeat_region   12084. .12929
                /rpt_family="L1MB2"
repeat_region   12930. .13224
                /rpt_family="AluSx"
repeat_region   13225. .13329
                /rpt_family="L1MB2"
repeat_region   13339. .13816
                /rpt_family="L1MA3"
repeat_region   13817. .14098
                /rpt_family="AluSp"
repeat_region   14099. .14124
                /rpt_family="(TAAA)n"
repeat_region   14125. .14281
                /rpt_family="L1MA3"
repeat_region   14806. .15181
                /rpt_family="MLT1A2"
repeat_region   complement(15305. .15344)
                /rpt_family="MIR"
repeat_region   complement(15345. .15695)
                /rpt_family="MLT1A1"
repeat_region   complement(15696. .15898)
                /rpt_family="MIR"
repeat_region   complement(16819. .17204)
                /rpt_family="Charlie1"
repeat_region   complement(17205. .17507)
                /rpt_family="AluY"
repeat_region   complement(17508. .18045)
                /rpt_family="Charlie1"
repeat_region   18047. .18196
                /rpt_family="AluSq"
repeat_region   18210. .18276
                /rpt_family="(TA)n"
repeat_region   complement(18306. .18346)
                /rpt_family="Charlie1"
repeat_region   18385. .18405
                /rpt_family="AT_rich"
repeat_region   complement(18407. .18702)
                /rpt_family="AluSq"
repeat_region   complement(18715. .18843)
                /rpt_family="Charlie1"
repeat_region   complement(18839. .18946)
                /rpt_family="Charlie1"
repeat_region   19305. .19595
                /rpt_family="AluSx"
repeat_region   23751. .23929
```

```
Alignment Scores:
Pred. No.:                1.96e+04        Length:         99545
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       8               Gaps:           0

US-09-858-332G-28 (1-6) x AC026115 (1-99545)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       5173 CATAATCATAATCATAAT 5156
```

```
RESULT 561
BX890595
LOCUS       BX890595               99914 bp    DNA     linear   VRT 27-OCT-2004
DEFINITION  Zebrafish DNA sequence from clone DKEY-104K17 in linkage group 5,
            complete sequence.
ACCESSION   BX890595
VERSION     BX890595.10  GI:54695487
KEYWORDS    HTG.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 99914)
  AUTHORS   Holt,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 27, 2004 this sequence version replaced gi:54311989.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived
            Zebrafish pUC subclones occasionally display inconsistency over the
            length of mononucleotide A/T runs and conserved TA repeats. Where
            this is found the longest good quality representation will be
            submitted.
            Repeat names beginning 'Dr' were identified by the Recon repeat
            discovery system (Zhirong Bao and Sean Eddy, submitted), and those
            beginning 'drr' were identified by Rick Waterman (Stephen Johnson
            lab, WashU).  For further information see
            http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml DKEY-104K17
```

```
           is from a Zebrafish BAC library
           VECTOR: pIndigoBAC-5.
FEATURES             Location/Qualifiers
     source          1. .99914
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-104K17"
                     /clone_lib="DanioKey"
ORIGIN

Alignment Scores:
Pred. No.:                1.97e+04       Length:         99914
Score:                    42.00          Matches:        6
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       5              Gaps:           0

US-09-858-332G-28 (1-6) x BX890595 (1-99914)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      52683 CATAATCATAATCATAAT 52700

RESULT 562
NCB23B10/c
LOCUS       NCB23B10               99967 bp    DNA     linear   PLN 17-APR-2005
DEFINITION  Neurospora crassa DNA linkage group II BAC contig B23B10.
ACCESSION   BX284752
VERSION     BX284752.1  GI:28881372
KEYWORDS    .
SOURCE      Neurospora crassa
  ORGANISM  Neurospora crassa
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
REFERENCE   1
  AUTHORS   Schulte,U., Aign,V., Hoheisel,J., Brandt,P., Fartmann,B.,
            Holland,R., Nyakatura,G., Mewes,H.W. and Mannhaupt,G.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 99967)
  AUTHORS   German Neurospora genome,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAR-2003) MIPS, Institut fuer Bioinformatik,
            GSF-Forschungszentrum fuer Umwelt und Gesundheit, GmbH,
            Ingolstaedter Landstrasse 1, D-85764 Neuherberg, FRG, E-mail:
            G.Mannhaupt@gsf.de Project Coordinator: Ulrich Schulte, Institute
            of Biochemistry, Heinrich-Heine-University, D-40225 Duesseldorf,
            E-mail: ulrich-schulte@uni-duesseldorf.de
COMMENT     this contig is an assembly of BAC 23B10 from 1 to: 41462, strain
            OR74A, cosmid 1K2 from 41463 to: 69790, cosmid library at DKFZ,
            strain OR74A and cosmid 39C11 from 69791 to 99967, cosmid library
            pLORIST6Xh, strain 74-OR-23-1A; BAC clone 23B10 and cosmid clone
            39C11 is available at the Fungal Genetic Stock Center,
            http://www.fgsc.net cosmid clone 1k2 is available from J. Hoheisel
            at DKFZ, e-mail:
            J.Hoheisel@dkfz-heidelberg.de
            Sequencing was performed by MWG Biotech AG, Ebersberg, Germany,
            http://www.mwgdna.com
            Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of linkage groups II
            and V can be viewed at: http://mips.gsf.de/proj/neurospora.
FEATURES             Location/Qualifiers
     source          1. .99967
                     /organism="Neurospora crassa"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5141"
                     /chromosome="6"
     gene            4533. .8183
                     /gene="B23B10.010"
     gene            complement(4533. .8183)
                     /gene="B23B10.010"
     CDS             complement(4533. .8183)
                     /gene="B23B10.010"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAD70414.1"
                     /db_xref="GI:28881373"
                     /db_xref="UniProt/TrEMBL:Q872Z6"
                     /translation="MYPHRRAPGDQGEHVIWRSNTADIPDDVIDPQLLAVSNQSDAAH
                     QPSLPQQPSQLLAETSQSNTVEQTSVYLHPPLDPVHQYNPQVELQPNRHVSLPPRDGS
                     IYQQQQLDFHHPNAFSNPHVNPQVISPSNSQPNPQYSFSPQSHSPFRDGSIHHHQPFN
                     PSSLAYSPPQDGRAFFQQAASSPSLPRHSRTNRMSQGDVRASPLIRTPHRNQDRSVRA
                     NVPSPAKSKSRAFSRGFGQGPMSEADEGPFRFQDEPDSSPLTSLTATPARSSSPDDGE
                     MEMDLDDWESGREAEVGIPSSSPPQNNVTPTRPRQRQEEDLQGREGQYQRPRRYNIDL
                     LLESVAITPVADMEIEEGVDEAASHADVESASHDPSLQRPESQENARPRAASLEIVTP
                     PRGSILRRFGPGTGLSFPSGVPYSTRYSDYTPPPEDFWRPLFAARPVPPGGYLAVERA
                     ADERTQTDNVSGPQDGESFRHNQHLNSQTGGVRHHMRTAIPGPSPTASTAPNNPQPTT
                     SRHGPASLPQGASQDTSVNDVVEVTSTDTILTHHRGPQRHQGHLRPLAPKPPIEDKDT
                     SPQPPHHTAQHPTQGNDQTQTQQPRFTNMAQFQPLLALPPPIQHPLLLSPFPPRIPGT
                     LTPALLSLAINPPRSLQLGTTDSLQPYLRACGLPVLYNDVYPRALPISVRREPFNRTV
                     IVEERYLAVIGLLYPRAEFIVSRIAEGVGNQKVEPNSGVMMDKWHVYKAVRMSGEEMQ
                     KLGVCVWSEKVCGEWQPVVEEREGREVKRVKMMDVVVDEEEEDETDEESDDIVVVSGR
                     ARASEKEGKENVRMTVGEERDEDCEDMEIVDAAAEGVQTSKTTNGKEKAKEVQVADKE
                     DKGKEILQEINDDDNTAAPPEPTSSSASAPLRRSTRVQSRQASVEEPATDKSENTTTA
                     PSNPSGKTLITRKLRAIQLDETGATITPNPNPTRPTKTPKPRLSTAITTTTTTTTKPP
                     APTVPFSNPSTVPSRSSNPTPSHRTSPTHYTVTLRQRDYPLGGVARIQCTPRARQVIE
                     SIQASFGVQHPYEVWRHKDYTVHDPDTGEVTQRYIKVLRGYAGDENTDDGLMRLFLDE
                     MDQRIKLKDTENKRNEKGKKKKDGQEEEEGKDMAALEDRVAWKERYLALKRRGKMAVA
                     RAKERKRDEEDMDMDYDEWLEERHEQEEELIEKWFAERSGLVEEEELARKTVEFALFE
                     YGGGDGLALGKP"
     exon            complement(4533. .8183)
                     /gene="B23B10.010"
                     /number=1
     tRNA            8734. .8806
                     /product="tRNA-Val"
                     /note="tRNA predict as a tRNA- Val : anticodon cac"
     gene            9902. .10129
                     /gene="B23B10.020"
     gene            complement(9902. .10129)
                     /gene="B23B10.020"
     CDS             complement(9902. .10129)
                     /gene="B23B10.020"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAD70415.1"
                     /db_xref="GI:28881374"
                     /db_xref="UniProt/TrEMBL:Q872Z5"
                     /translation="MSSSNTDTSTTTTSGDNTSSTSTTTQTNSSTTSSSSSSSTSTTS
                     KTTGSSGYSYIERLRIEGQAQYWLGPEKKSS"
     exon            complement(9902. .10129)
                     /gene="B23B10.020"
                     /number=1
     gene            13672. .15211
                     /gene="B23B10.030"
     CDS             join(13672. .14823,14894. .15211)
                     /gene="B23B10.030"
                     /note="similarity to monocarboxylate transporter (mct3),
                     homo sapiens, TREMBL:U81800_1
                     Contains Multicopper oxidases signatures
                     [GPYVLFTFGTFVGFIGLYVML]
                     contains EST gb:AA898910, BG279307"
                     /codon_start=1
                     /product="related to monocarboxylate transporter"
                     /protein_id="CAD70416.1"
                     /db_xref="GI:28881375"
                     /db_xref="GOA:Q872Z4"
                     /db_xref="InterPro:IPR002355"
                     /db_xref="InterPro:IPR007114"
                     /db_xref="InterPro:IPR011701"
                     /db_xref="UniProt/TrEMBL:Q872Z4"
                     /translation="MAPTTFPDRTTHGPSTDAEGNLSDQSTATGISSPSRASSEKSSP
                     HVTAAPPLTGAAASSVAPDAEKQQPPPTTAAWNAGPPPPNGGLAAWLQVVSVFFIFFN
                     TWGLLNTFGIFQTYYESSESIFQPPGSVSSSNISWIGSIQSFMVLLGGLVSGPLYDRG
                     YLKYLIYVGAFGVVFGHMMLSLVSEFWQALLAQGFVVGLGAGMMFTPGVTVLQSYFST
                     KIGLATGLAAAGSSFGGVIYPIVFYRLIDRIGYGWSVRVIGFIALATLMVPVVFLKQR
```

```
                TKSGKVRALVDVSAFTDGPYVLFTFGTFVGFIGLYVMLFYLSFFGLDTGATDPAMAFY
                IVPILNAASMFGRTLPNWLSDKVGPFNILIPGALVCGALTLAMMGVKGLGAIIVVAIL
                FGPFSGVFIALPTVCYIALTKDKSRLGSRIGLGYAFLGLATLIGGPGTGAILQQTGHN
                GHNNWTGIWVFGGVMLLAASGIFMVLRSWKFGFKLTTKA"
     exon       13672. .14823
                /gene="B23B10.030"
                /number=1
     intron     14824. .14893
                /gene="B23B10.030"
                /number=1
     exon       14894. .15211
                /gene="B23B10.030"
                /number=2
     gene       19788. .21920
                /gene="B23B10.040"
     gene       complement(join(19788. .21110,21174. .21622,21848. .21920))
                /gene="B23B10.040"
     CDS        complement(join(19788. .21110,21174. .21622,21848. .21920))
                /gene="B23B10.040"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="CAD70417.1"
                /db_xref="GI:28881376"
                /db_xref="GOA:Q872Z3"
                /db_xref="InterPro:IPR002156"
                /db_xref="UniProt/TrEMBL:Q872Z3"
                /translation="MDKGQPPDHPASEQEPPAKRHKPLNSYFPQPFRGKIFVQNERPR
                EEEAIAAAREAIASAKRSTENDFTIAFFTDASVEEKTPDLPGGYSVVFDNFFPGTHHY
                RERLGMAWLMPNMITSTIGEAVGILQALQVAIHHLRAIAISPADTASARRVTLCIFSD
                AHQVIRLIKCQNLNVGDIPGFDVFLELHWVKGHAGVEENDQADLIARRARPIEDGRDI
                FVASGADVDATALPPAVHLLFLDDIAAEKARQRPPRPLGKQNDRRKRNGKRAHRRRDE
                NVPSSSLHTPNDDMTYSPPPWASDSFRFSTNSLEMPQYPATRPFSFASEEHPRVSEKP
                TLETEEQSPMPKQYTPLDVNNHEKKPAVEIEALERDSTLEQGVAVIQPAVIQQAIVEK
                EAAVLNINKPTIGLDSSANEQALTLLQHEEQEELMTDLEVQVPEKNSVLEHIVSDSLA
                SVPQQVVEEDKHSVKREVQKPAILLSTGMEVAENKQDATPHSVNEQHEPSTKAEDQKF
                IIPEAIEDTEPSPDSEGSTAVHQPAEEHNDSALEPLLGDQALIDLQLEGEYVEWCQRR
                LQDVLVSVYNHEGGSTGETDTSEEVISGEETQAAVGRPLAEEVEATLA"
     exon       complement(19788. .21110)
                /gene="B23B10.040"
                /number=1
     intron     complement(21111. .21173)
                /gene="B23B10.040"
                /number=1
     exon       complement(21174. .21622)
                /gene="B23B10.040"
                /number=2
     intron     complement(21623. .21847)
                /gene="B23B10.040"
                /number=2
     exon       complement(21848. .21920)
                /gene="B23B10.040"
                /number=3
     gene       22992. .23446
                /gene="B23B10.050"
     gene       complement(join(22992. .23008,23188. .23446))
                /gene="B23B10.050"
     CDS        complement(join(22992. .23008,23188. .23446))
                /gene="B23B10.050"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="CAD70418.1"
                /db_xref="GI:28881377"
                /db_xref="UniProt/TrEMBL:Q872Z2"
                /translation="MSNIFFWLPMEDQAKVYALMWPGNWKKKKKKKKKKKKKKKKSSA
                SLASQSMGASGSGQKEKASSDGGSGGSKDGKMETEVETVEYKESWAR"
     exon       complement(22992. .23008)
                /gene="B23B10.050"
                /number=1
     intron     complement(23009. .23187)
                /gene="B23B10.050"
                /number=1
     exon       complement(23188. .23446)
                /gene="B23B10.050"
                /number=2
     repeat_region  23323. .23370
                /note="48 bp tcttcttcttcttcttcttcttcttcttcttcttct tandem
                repeat"
     gene       24423. .26455
                /gene="B23B10.060"
     CDS        join(24423. .24497,24557. .26455)
                /gene="B23B10.060"

Alignment Scores:
Pred. No.:              1.97e+04        Length:         99967
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x NCB23B10 (1-99967)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       5304 CATAACCACAACCACAAC 5287

RESULT 563
AP000093
LOCUS       AP000093              100000 bp    DNA     linear   PRI 25-MAY-2002
DEFINITION  Homo sapiens genomic DNA of 21q22.1, GART and AML related,
            B335D16-P10G11 region, segment 3/7.
ACCESSION   AP000093
VERSION     AP000093.1  GI:4730854
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
  TITLE     Homo sapiens 726,344bp genomic DNA of 21q22.1 GART and AML region
  JOURNAL   Published Only in Database (1999)
REFERENCE   2  (bases 1 to 100000)
  AUTHORS   Hirakawa,M., Yamaguchi,H., Imai,K. and Shimada,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-APR-1999) Mika Hirakawa, Japan Science and Technology
            Corporation (JST), Advanced Databases Department; 5-3, Yonbancho,
            Chiyoda-ku, Tokyo 102-0081, Japan (E-mail:mika@tokyo.jst.go.jp,
            URL:http://www-alis.tokyo.jst.go.jp/, Tel:81-3-5214-8491,
            Fax:81-3-5214-8470)
COMMENT     This sequence is conducted by Kitasato University JST sequencing
            Laboratory as a JST sequencing team.
            Principal Investigator:Yoshiyuki Sakaki Ph.D.
            Phone: +81-3-5449-5622, Fax : +81-3-5449-5445,
            sakaki@hgc.ims.u-tokyo.ac.jp
            Sub-leader: Tadayoshi Shiba Ph.D., Masahira Hattori Ph.D. The
            sequence is submitted by Human Genome Sequencing in ALIS project of
            JST.
            Japan Science and Technology Corporation (JST)
            5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081 Japan
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www-alis.tokyo.jst.go.jp/HGS/) or send
            email to webmaster@www-alis.tokyo.jst.go.jp.
FEATURES              Location/Qualifiers
     source           1. .100000
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /chromosome="21"
                      /map="21q22.1"
     STS              26117. .26377
                      /standard_name="D21S300"
                      /note="G52E07;The location is between each flanking site
                      of PCR primers."
```

```
                     /db_xref="GDB:190673"
ORIGIN

Alignment Scores:
Pred. No.:                1.97e+04        Length:         100000
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       8               Gaps:           0

US-09-858-332G-28 (1-6) x AP000093 (1-100000)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      64928 CATAATCATAATCATAAT 64945

RESULT 564
AP000197
LOCUS       AP000197              100000 bp    DNA     linear   PRI 08-JAN-2000
DEFINITION  Homo sapiens genomic DNA, chromosome 21q22.1, D21S226-AML region,
            clone B355D16-T1073, segment 3/8, complete sequence.
ACCESSION   AP000197
VERSION     AP000197.1  GI:4827095
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 100000)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.
  TITLE     Homo sapiens 726,344bp genomic DNA of 21q22.1 (REGION: D21S226-AML
            CLONE RANGE: B355D16-T1073)
  JOURNAL   Published Only in DataBase (1999)
REFERENCE   2  (bases 1 to 100000)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-1999) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     E. coli transposon insertion:The present data does not contain E.
            coli transposon sequences which integrated in the
            original/previous sequences.  We determined the boundary between
            the insertion and genomic sequences experimentally, removed the
            insertion sequences, reconstituted the present data. The sequencing
            project is supported by Japan Science Technology Corporation (JST)
            and The Institute of Physical and Chemical Research (RIKEN).
FEATURES             Location/Qualifiers
     source          1. .100000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
ORIGIN

Alignment Scores:
Pred. No.:                1.97e+04        Length:         100000
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       8               Gaps:           0

US-09-858-332G-28 (1-6) x AP000197 (1-100000)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      64928 CATAATCATAATCATAAT 64945

RESULT 565
AC161017/c
LOCUS       AC161017              101021 bp    DNA     linear   HTG 05-MAY-2005
DEFINITION  Pan troglodytes chromosome Y clone CH251-157N1, WORKING DRAFT
            SEQUENCE, 8 unordered pieces.
ACCESSION   AC161017
VERSION     AC161017.1  GI:63028373
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pan.
REFERENCE   1  (bases 1 to 101021)
  AUTHORS   Wilson,R.K.
  TITLE     The sequence of Pan troglodytes clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 101021)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAY-2005) Genetics, Genome Sequencing Center, 4444
            Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT
            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu
            Contact: submissions@watson.wustl.edu
            ---------- Project Information ----------
            Center project name: C_AB0157N01
            ----------
            ---------- Summary Statistics  ----------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 99195 bases at least Q40
            Consensus quality: 99387 bases at least Q30
            Consensus quality: 99605 bases at least Q20
            ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1227: contig of 1227 bp in length
            *     1228     1327: gap of unknown length
            *     1328     2900: contig of 1573 bp in length
            *     2901     3000: gap of unknown length
            *     3001     4736: contig of 1736 bp in length
            *     4737     4836: gap of unknown length
            *     4837     9906: contig of 5070 bp in length
            *     9907    10006: gap of unknown length
            *    10007    17896: contig of 7890 bp in length
            *    17897    17996: gap of unknown length
            *    17997    40486: contig of 22490 bp in length
            *    40487    40586: gap of unknown length
            *    40587    99657: contig of 59071 bp in length
            *    99658    99757: gap of unknown length
            *    99758   101021: contig of 1264 bp in length.
FEATURES             Location/Qualifiers
     source          1. .101021
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /chromosome="Y"
```

```
                /clone="CH251-157N1"
     misc_feature    1. .1227
                /note="assembly_name:Contig17"
     gap             1228. .1327
                /estimated_length=unknown
     misc_feature    1328. .2900
                /note="assembly_name:Contig18"
     gap             2901. .3000
                /estimated_length=unknown
     misc_feature    3001. .4736
                /note="assembly_name:Contig20"
     gap             4737. .4836
                /estimated_length=unknown
     misc_feature    4837. .9906
                /note="assembly_name:Contig21"
     gap             9907. .10006
                /estimated_length=unknown
     misc_feature    10007. .17896
                /note="assembly_name:Contig22"
     gap             17897. .17996
                /estimated_length=unknown
     misc_feature    17997. .40486
                /note="assembly_name:Contig23"
     gap             40487. .40586
                /estimated_length=unknown
     misc_feature    40587. .99657
                /note="assembly_name:Contig24
                clone_end:SP6
                vector_side:right"
     gap             99658. .99757
                /estimated_length=unknown
     misc_feature    99758. .101021
                /note="assembly_name:Contig5"
ORIGIN

Alignment Scores:
Pred. No.:              1.99e+04        Length:         101021
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AC161017 (1-101021)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      33712 CATAATCATAATCATAAT 33695

RESULT 566
CNS07TIW/c
LOCUS       CNS07TIW              101256 bp    DNA     linear   PLN 24-OCT-2001
DEFINITION  DNA centromeric region sequence from BAC DP35C12 of chromosome 5 of
            Podospora anserina.
ACCESSION   AL627322
VERSION     AL627322.1  GI:16444879
KEYWORDS    .
SOURCE      Podospora anserina
  ORGANISM  Podospora anserina
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetidae; Sordariales; Lasiosphaeriaceae; Podospora.
REFERENCE   1  (bases 1 to 101256)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT-2001) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     * Christian Barreau and Beatrice Turcq, Institut de Biochimie et de
            Genetique Cellulaire, UPR CNRS 9026, 1 rue Camille Saint Saens,
            F-33077 Bordeaux, France;
            * Alain Billault, Molecular Engines Laboratories, 20 rue Bouvier,
            F-75011 Paris, France;
            * Laurence Cattolico and Simone Duprat, Genoscope, Centre National
            de Sequencage, 2 rue Gaston Cremieux, CP5706, F-91057 Evry CEDEX,
            France; * Robert Debuchy, Sebastien Kicka and Philippe Silar,
            Institut de Genetique et Microbiologie, UMR CNRS 8621, batiment
            400, Universite Paris-Sud, F-91405 Orsay CEDEX, France;
            * Annie Sainsard-Chanet and Carole H. Sellem, Centre de Genetique
            Moleculaire, UPR CNRS 9061, batiment 24, allee de la terrasse,
            F-91190 Gif sur Yvette, France;
            base calling/assembly program : Phred/Phrap;
            consensus quality: 98,93% of bases at least Q40.
FEATURES            Location/Qualifiers
     source         1. .101256
                    /organism="Podospora anserina"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:5145"
                    /chromosome="5"
                    /clone="DP35C12"
ORIGIN

Alignment Scores:
Pred. No.:              2e+04           Length:         101256
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x CNS07TIW (1-101256)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      49868 CACAACCACAACCACAAC 49851

RESULT 567
AC102942
LOCUS       AC102942              101620 bp    DNA     linear   PRI 13-MAY-2002
DEFINITION  Homo sapiens chromosome 8, clone RP11-237O13, complete sequence.
ACCESSION   AC102942
VERSION     AC102942.2  GI:20531846
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 101620)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-237O13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 101620)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
```

```
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 101620)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 13, 2002 this sequence version replaced gi:17061052.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L20298
                Center clone name: 237_O_13
            --------------

            Only the first 101.6 kilobases of this clone are being submitted.
            The remainder overlaps accession number AC098612 [Jena project].
FEATURES             Location/Qualifiers
     source          1. .101620
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-237O13"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   9. .128
                     /rpt_family="L3"
     repeat_region   complement(539. .849)
                     /rpt_family="AluJo"
     repeat_region   complement(1005. .1471)
                     /rpt_family="LTR40b"
     repeat_region   complement(1500. .1768)
                     /rpt_family="L1PA8"
     repeat_region   2244. .2383
                     /rpt_family="(TAGA)n"
     repeat_region   complement(2384. .2514)
                     /rpt_family="MER5A"
     repeat_region   3439. .3487
                     /rpt_family="AT_rich"
     repeat_region   complement(3776. .4073)
                     /rpt_family="L1ME2"
     repeat_region   4095. .4403
                     /rpt_family="MLT2D"
     repeat_region   5098. .5119
                     /rpt_family="AT_rich"
     repeat_region   complement(5237. .5632)
                     /rpt_family="L2"
     repeat_region   5758. .5914
                     /rpt_family="L1ME1"
     repeat_region   complement(5925. .6949)
                     /rpt_family="L1"
     repeat_region   complement(6952. .7213)
                     /rpt_family="L1"
     repeat_region   complement(7229. .8355)
                     /rpt_family="L1"
     repeat_region   complement(8356. .8672)
                     /rpt_family="AluSx"
     repeat_region   complement(8673. .10154)
                     /rpt_family="L1"
     unsure          8986. .8998
                     /note="Some clones have 12, not 13, T's"
     repeat_region   10680. .10717
                     /rpt_family="L1ME1"
     repeat_region   10718. .11029
                     /rpt_family="AluSx"
     repeat_region   11030. .11210
                     /rpt_family="L1ME1"
     repeat_region   11212. .11240
                     /rpt_family="AT_rich"
     repeat_region   11253. .11443
                     /rpt_family="L1PBa"
     repeat_region   11446. .12795
                     /rpt_family="L1PBa"
     repeat_region   12800. .13859
                     /rpt_family="L1PB2"
     repeat_region   13836. .14308
                     /rpt_family="L1ME1"
     repeat_region   14348. .14436
                     /rpt_family="MLT1E1"
     repeat_region   14503. .15576
                     /rpt_family="L1ME1"
     repeat_region   15604. .16365
                     /rpt_family="L1M4c"
     repeat_region   16352. .17422
                     /rpt_family="L1PA15"
     repeat_region   17426. .17680
                     /rpt_family="L1M4c"
     repeat_region   17698. .17960
                     /rpt_family="L1M4c"
     repeat_region   17961. .18182
                     /rpt_family="AluJb"
     repeat_region   18194. .18243
                     /rpt_family="GA-rich"
     repeat_region   18251. .18319
                     /rpt_family="L1M4c"
     repeat_region   complement(18320. .18994)
                     /rpt_family="LTR12"
     repeat_region   complement(18995. .19046)
                     /rpt_family="LTR30"
     repeat_region   19047. .19246
                     /rpt_family="L1M4c"
     repeat_region   19223. .19358
                     /rpt_family="L1MB8"
     repeat_region   19357. .19852
                     /rpt_family="THE1B-int"
     repeat_region   19853. .20195
                     /rpt_family="THE1B"
     repeat_region   complement(20197. .20308)
                     /rpt_family="L2"
     repeat_region   complement(20351. .20523)
                     /rpt_family="MIR"
     repeat_region   complement(20557. .20658)
                     /rpt_family="MIR"
     repeat_region   20768. .20792
                     /rpt_family="(T)n"
     repeat_region   20803. .21005
                     /rpt_family="MIR"
     repeat_region   21155. .21189
                     /rpt_family="(CA)n"
```

```
     repeat_region   21328. .21396
                     /rpt_family="L2"
     repeat_region   complement(22428. .22548)
                     /rpt_family="MER5B"
     repeat_region   complement(22549. .22862)
                     /rpt_family="AluSg"
     repeat_region   complement(22863. .22921)
                     /rpt_family="MER5B"
     repeat_region   23012. .23264
                     /rpt_family="LTR33A"
     repeat_region   23265. .23630
                     /rpt_family="L2"
     repeat_region   complement(23631. .23920)
                     /rpt_family="AluSx"
     repeat_region   23933. .24038

Alignment Scores:
Pred. No.:              2e+04          Length:         101620
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     8              Gaps:           0

US-09-858-332G-28 (1-6) x AC102942 (1-101620)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      43273 CATAATCATAACCATAAT 43290

RESULT 568
AP006378
LOCUS       AP006378              101752 bp    DNA     linear   PLN 22-JUL-2003
DEFINITION  Lotus corniculatus var. japonicus genomic DNA, chromosome 1,
            clone:LjT44A10, TM0223, complete sequence.
ACCESSION   AP006378
VERSION     AP006378.1  GI:31581009
KEYWORDS    HTG.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1
  AUTHORS   Asamizu,E., Kato,T., Sato,S., Nakamura,Y., Kaneko,T. and Tabata,S.
  TITLE     Structural Analysis of a Lotus japonicus Genome. IV. Sequence
            Features and Mapping of seventy-three TAC clones which cover the
            7.5 Mb Regions of the Genome
  JOURNAL   DNA Res. (2003) In press
REFERENCE   2  (bases 1 to 101752)
  AUTHORS   Sato,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAY-2003) Shusei Sato, Kazusa DNA Research Institute,
            Department of Plant Gene Research; 2-6-7 Kazusa-kamatari, Kisarazu,
            Chiba 292-0818, Japan (E-mail:ssato@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/, Tel:81-438-52-3935(ex.2337),
            Fax:81-438-52-3934)
FEATURES             Location/Qualifiers
     source          1. .101752
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="genomic DNA"
                     /variety="japonicus"
                     /db_xref="taxon:34305"
                     /chromosome="1"
                     /clone="LjT44A10"
                     /clone_lib="LjT library"
                     /note="TAC clone:TM0223
                     synonym: Lotus japonicus"
ORIGIN

Alignment Scores:
Pred. No.:              2.01e+04       Length:         101752
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     15             Gaps:           0

US-09-858-332G-28 (1-6) x AP006378 (1-101752)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      33052 CACAATCACAATCACAAT 33069

RESULT 569
AC005395/c
LOCUS       AC005395              102004 bp    DNA     linear   PLN 11-MAR-2002
DEFINITION  Arabidopsis thaliana chromosome 2 clone F17H15 map B68, complete
            sequence.
ACCESSION   AC005395
VERSION     AC005395.3  GI:20197358
KEYWORDS    HTG.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 102004)
  AUTHORS   Rounsley,S.D., Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Mason,T.M.,
            Shen,M., Ronning,C.M., Fraser,C.M., Somerville,C.R. and Venter,J.C.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 102004)
  AUTHORS   Lin,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 102004)
  AUTHORS   Town,C.D. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
COMMENT     On Apr 18, 2002 this sequence version replaced gi:6598463.
FEATURES             Location/Qualifiers
     source          1. .102004
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /chromosome="2"
                     /map="B68"
                     /clone="F17H15"
                     /ecotype="Columbia"
     misc_feature    1. .18556
                     /note="overlap with BAC clone T19L18
                     (AC004747:68524. .87080)."
     gene            368. .4128
                     /gene="At2g25970"
                     /note="synonym: F17H15.1"
     mRNA            join(<368. .701,796. .1079,1169. .1270,1343. .1414,1500. .1619,
                     3142. .>4128)
                     /gene="At2g25970"
     CDS             join(368. .701,796. .1079,1169. .1270,1343. .1414,1500. .1619,
                     3142. .4128)
                     /gene="At2g25970"
                     /codon_start=1
                     /product="unknown protein"
                     /protein_id="AAM15045.1"
                     /db_xref="GI:20197368"
                     /translation="MADESQYSSDTYSNKRKYEEPTAPPPSTRRPTGFSSGPIPSASV
                     DPTAPTGLPPSSYNSVPPPMDEIQIAKQKAQEIAARLLNSADAKRPRVDNGASYDYGD
                     NKGFSSYPSEGKQMSGTVPSSIPVSYGSFQGTTKKIDIPNMRVGVIIGKGGETIKYLQ
                     LQSGAKIQVTRDMDADPNCATRTVDLTGTPDQISKAEQLITDVLQEAEAGNTAGSGGG
                     GGRRMGGQAGADQFVMKIPNNKVGLIIGKGGETIKSMQAKTGARIQVIPLHLPPGDPT
                     PERTLQIDGITEQIEHAKQLVNEIISGENRMRNSAMGGGYPQQGGYQARPPSSWAPPG
                     GPPAQPGYGGYMQPGAYPGPPQYGQSPYGSYPQQTSAGYYDQSSVPPSQQSAQGEYDY
                     YGQQQSQQPSSGGSSAPPTDTTGYNYYQHASGYGQAGQGYQQDGYGAYNASQQSGYGQ
```

AAGYDQQGGYGSTTNPSQEEDASQAAPPSSAQSGQAGYGTTGQQPPAQGSTGQAGYGA
PPTSQAGYSSQPAAAYNSGYGAPPPASKPPTYGQSQQSPGAPGSYGSQSGYAQPAASG
YGQPPAYGYGQAPQGYGSYGGYTQPAAGGGYSSDGSAGATAGGGGGTPASQSAAPPAG
PPKASPKS"

gene complement(5312. .7321)
/gene="At2g25950"
/note="synonym: F17H15.2; supported by full length cDNA:
Ceres:39666"

mRNA complement(join(<5312. .5490,5570. .5631,5712. .5820,
5998. .6047,6134. .6191,6302. .6379,6566. .6615,6785. .6865,
7154. .>7321))
/gene="At2g25950"

CDS complement(join(5448. .5490,5570. .5631,5712. .5820,
5998. .6047,6134. .6191,6302. .6379,6566. .6615,6785. .6865,
7154. .7237))
/gene="At2g25950"
/codon_start=1
/product="expressed protein"
/protein_id="AAM15044.1"
/db_xref="GI:20197367"
/translation="MSCTHDHNCEDHECSSDWSLYKHIDLSKVSALNESVSGSAKSVF
KAWEQRLHSSGEHLESNEGDPELLVFVPFTSDVKIKSISIVGGPEGTSPSKLRVFINR
EGIDFSDAESMQAVQEWELAENLQGVLEYQTRYSKFQSVGNITLHFPESFGGDTTQIR
YIGFKGEATQLKRDVVATIVYEIRPNPSDHKTKAETGGGFSQVE"

repeat_region complement(8934. .8974)
/rpt_family="AT_rich"

repeat_region complement(9135. .9157)
/rpt_family="AT_rich"

repeat_region complement(9616. .9637)
/rpt_family="AT_rich"

gene 9944. .12647
/gene="At2g25940"
/note="synonym: F17H15.4; identical to GB:D61393;
supported by cDNA: gi_16604313_gb_AY058055 1_"

mRNA join(9944. .10134,10550. .10714,10794. .10950,11037. .11122,
11197. .11445,11560. .11769,11865. .12068,12294. .12647)
/gene="At2g25940"

CDS join(9952. .10134,10550. .10714,10794. .10950,11037. .11122,
11197. .11445,11560. .11769,11865. .12068,12294. .12476)
/gene="At2g25940"
/codon_start=1
/product="putative vacuolar processing enzyme"
/protein_id="AAM15043.1"
/db_xref="GI:20197366"
/translation="MTTVVSFLALFLFLVAAVSGDVIKLPSLASKFFRPTENDDDSTK
WAVLVAGSSGYWNYRHQADVCHAYQLLKKGGVKEENIVVFMYDDIAKNEENPRPGVII
NSPNGEDVYNGVPKDYTGDEVNVDNLLAVILGNKTALKGGSGKVVDSGPNDHIFIYYS
DHGGPGVLGMPTSPNLYANDLNDVLKKKYASGTYKSLVFYLEACESGSIFEGLLPEGL
NIYATTASNAEESSWGTYCPGEDPSPPSEYETCLGDLYSVAWIEDSEKHNLQTETLHE
QYELVKKRTAGSGKSYGSHVMEFGDIGLSKEKLVLFMGTNPADENFTFVNENSIRPPS
RVTNQRDADLVHFWHKYQKAPEGSARKVEAQKQVLEAMSHRLHVDNSILLIGILLFGL
EGHAVLNKVRPSGEPLVDDWDCLKSLVRAFERHCGSLSQYGIKHMRSIANMCNAGIQM
RQMEEAAMQACPTIPTSPWSSLDRGFSA"

gene complement(12794. .16513)
/gene="At2g25930"
/note="synonym: F17H15.25"

mRNA complement(join(<12794. .13813,13904. .13955,15136. .15932,
16295. .>16513))
/gene="At2g25930"

CDS complement(join(12794. .13813,13904. .13955,15136. .15932,
16295. .16513))
/gene="At2g25930"
/codon_start=1
/product="unknown protein"
/protein_id="AAM15042.1"
/db_xref="GI:20197364"
/translation="MKRGKDEEKILEPMFPRLHVNDADKGGPRAPPRNKMALYEQLSI
PSQRFGDHGTMNSRSNNTSTLVHPGPSSQPCGVERNLSVQHLDSSAANQATEKFVSQM
SFMENVRSSAQHDQRKMVREEEDFAVPVYINSRRSQSHGRTKSGIEKEKHTPMVAPSS
HHSIRFQEVNQTGSKQNVCLATCSKPEVRDQVKANARSGGFVISLDVSVTEEIDLEKS
ASSHDRVNDYNASLRQESRNRLYRDGGKTRLKDTDNGAESHLATENHSQEGHGSPEDI
DNDREYSKSRACASLQQINEEASDDVSDDSMVDSISSIDVSPDDVVGILGQKRFWRAR
KAIANQQRVFAVQLFELHRLIKVQKLIAASPDLLLDEISFLGKVSAKSYPVKKLLPSE
FLVKPPLPHVVVKQRGDSEKTDQHKMESSAENVVGRLSNQGHHQQSNYMPFANNPPAS
PAPNGYCFPPQPPPSGNHQQWLIPVMSPSEGLIYKPHPGMAHTGHYGGYYGHYMPTPM
VMPQYHPGMGFPPPGNGYFPPYGMMPTIMNPYCSSQQQQQQQPNEQMNQFGHPGNLQN
TQQQQQRSDNEPAPQQQQQPTKSYPRARKSRQGSTGSSPSGPQGISGSKSFRPFAAVD
EDSNINNAPEQTMTTTTTTTRTTVTQTTRDGGGVTRVIKVVPHNAKLASENAARIFQS
IQEERKRYDSSKP"

repeat_region complement(13231. .13252)
/rpt_family="(CAA)n"

repeat_region 19162. .19183
/rpt_family="AT_rich"

repeat_region complement(19854. .19890)
/rpt_family="AT_rich"

gene 19930. .21718
/gene="At2g25920"
/note="synonym: F17H15.5; predicted by genscan; supported
by cDNA: gi_15294291_gb_AF410337 1_AF410337"

mRNA join(19930. .20110,20566. .20705,20796. .20893,20998. .21718)
/gene="At2g25920"

CDS join(20056. .20110,20566. .20705,20796. .20893,20998. .21547)
/gene="At2g25920"
/codon_start=1
/product="expressed protein"
/protein_id="AAC42240.2"
/db_xref="GI:20197362"
/translation="MTSSSPTQRTHVSIPPEPGGKSLTQEANEPPVPISVSSPCGKRT
RDPEDEVYLDNLRSQKRYLSEIMACSLNGLTVGDSLPVNMLESPARSESFLYHRDDLS
LQYSPMSEDSDEARFCEDPTATASTSSSQPESRPTSPVSPYRYQRPLTSTNSPQPSPT
ILHHSHTCPASMISNAATTTTPQSRQRGSDTEGRFPSSPSDICHSGDLRRTALLRSVQ
MRTQPCGYSSSSGPSNIDGEERMCSKSMEEDRGYNKGEDIPYAEVSSKSKSCKALDMR
LCEK"

repeat_region complement(21596. .21666)
/rpt_family="AT_rich"

repeat_region complement(21680. .21702)
/rpt_family="AT_rich"

repeat_region 21963. .22002
/rpt_family="AT_rich"

repeat_region 22267. .22343
/rpt_family="AT_rich"

repeat_region complement(22409. .22434)
/rpt_family="AT_rich"

gene 23965. .26908
/gene="At2g25910"
/note="synonym: F17H15.6; supported by full length cDNA:
Ceres:26655"

mRNA join(23965. .24029,24558. .24733,24818. .24949,25037. .25087,
25184. .25280,25536. .25582,25690. .25909,26128. .26257,
26350. .26392,26519. .26908)
/gene="At2g25910"

CDS join(23975. .24029,24558. .24733,24818. .24949,25037. .25087,
25184. .25280,25536. .25582,25690. .25909,26128. .26257,
26350. .26392,26519. .26593)
/gene="At2g25910"
/codon_start=1
/product="expressed protein"
/protein_id="AAC42241.2"
/db_xref="GI:20197359"
/translation="MASSSPTQLAHVPIPPEPGGRSPTQEANEPPVPIYIVTDPFQLP
ADFLNPSPEKKLVIGFDCEGVDLCRHGKLCIMQIAFSNAIYLVDVIEGGEVIMKACKP
ALESNYITKVIHDCKRDSEALYFQFGIRLHNVVDTQIAYSLIEEQEGRRRPLDDYISF
VSLLADPRYCGISYEEKEEVRVLMRQDPKFWTYRPMTELMIRAAADDVRFLLYLYHKM
MGKLNQRSLWHLAVRGALYCRCLCCMNDADFADWPTVPPIPDNLKSEDQCLEEEILSV
LDVPPGKMGRVIGRKGASILAIKEACNAEILIGGAKGPPDKIFVIGPVREVRKAEAIL
RGRMIDY"

repeat_region 27446. .27473
/rpt_family="AT_rich"

repeat_region complement(27866. .27926)
/rpt_family="(TAAAAA)n"

repeat_region complement(30370. .30442)

Alignment Scores:

Pred. No.:	2.01e+04	Length:	102004
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0

```
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     15            Gaps:          0

US-09-858-332G-28 (1-6) x AC005395 (1-102004)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      23252 CATAATCATAACCATAAC 23235

RESULT 570
BX000688
LOCUS       BX000688              102684 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone DAQB-36F16 on chromosome 6 Contains
            the GPR53P pseudogene for G protein-coupled receptor 53, the OR2I1P
            pseudogene for olfactory receptor, family 2, subfamily I, member 1,
            the UBD gene for ubiquitin D, the OR2H5P pseudogene for olfactory
            receptor, family 2, subfamily H, member 5, the RPL13AP pseudogene
            for ribosomal protein L13a, the OR2H2 gene for olfactory receptor,
            family 2, subfamily H, member 2, the GABBR1 gene for
            gamma-aminobutyric acid (GABA) B receptor, 1, and 3 CpG islands,
            complete sequence.
ACCESSION   BX000688
VERSION     BX000688.11  GI:32187946
KEYWORDS    HTG; CpG island; GABBR1; GPR53P; OR2H2; OR2H5P; OR2I1P; RPL13AP;
            UBD.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 102684)
  AUTHORS   Almeida,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 24, 2003 this sequence version replaced gi:32168690.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs constructed by
            the MHC Haplotype Consortium and collaborators. Further information
            can be found at
            http://www.sanger.ac.uk/HGP/Chr6/MHC
            DAQB-36F16 is from a DNA-arts QBL human bac library VECTOR:
            pBeloBAC11
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.
FEATURES             Location/Qualifiers
     source          1. .102684
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="DAQB-36F16"
                     /clone_lib="DNA-arts-BAC.1-QBL.1"
     misc_feature    1
                     /note="Clone_left_end: DAQB-36F16"
     polyA_signal    1202. .1207
     polyA_site      1220
     gene            complement(5425. .6214)
                     /locus_tag="DAQB-36F16.6-001"
                     /pseudo
     CDS             complement(5425. .6214)
                     /locus_tag="DAQB-36F16.6-001"
                     /note="match: proteins: P35410 Q91ZB7 Q99MT8"
                     /pseudo
                     /codon_start=1
                     /product="G protein-coupled receptor 53, pseudogene"
     gene            20913. .21845
                     /locus_tag="DAQB-36F16.4-001"
                     /pseudo
     CDS             20913. .21845
                     /locus_tag="DAQB-36F16.4-001"
                     /note="match: proteins: Q8N349 Q8NGS6 Q8VGR4 Q8VGT6"
                     /pseudo
                     /codon_start=1
                     /product="olfactory receptor, family 2, subfamily I,
                     member 1 pseudogene"
     gene            complement(23299. .27606)
                     /gene="UBD"
                     /locus_tag="DAQB-36F16.5-001"
     mRNA            complement(join(23299. .24030,27356. .27606))
                     /gene="UBD"
                     /locus_tag="DAQB-36F16.5-001"
                     /product="ubiquitin D"
                     /note="match: ESTs: BI760949.1 CB985708.1
                     match: cDNAs: AF123050.1 BC012472.1 Y12653.1"
     polyA_site      complement(23305)
                     /gene="UBD"
                     /locus_tag="DAQB-36F16.5-001"
     polyA_signal    complement(23323. .23328)
                     /gene="UBD"
                     /locus_tag="DAQB-36F16.5-001"
     CDS             complement(join(23560. .24030,27356. .27382))
                     /gene="UBD"
                     /locus_tag="DAQB-36F16.5-001"
                     /standard_name="OTTHUMP00000014898"
                     /note="match: proteins: O15205 P93501 Q96EC7"
                     /codon_start=1
                     /product="ubiquitin D"
                     /protein_id="CAI18599.1"
                     /db_xref="GI:56207879"
                     /db_xref="GOA:Q5RJ89"
                     /db_xref="InterPro:IPR000626"
                     /db_xref="UniProt/TrEMBL:Q5RJ89"
                     /translation="MAPNASCLCVHVRSEEWDLMTFDANPYDSVKKIKEHVRSKTKVP
                     VQDQVLLLGSKILKPRRSLSSYGTDKEKTIHLTLKVVKPSDEELPLFLVESGDEAKRH
                     LLQVRRSSSVAQVKAMIETKTGIIPETQIVTCNGKRLEDGKMMADYGIRKGNLLFLAS
                     YCIGG"
     gene            41602. .42529
                     /locus_tag="DAQB-36F16.1-001"
                     /pseudo
     CDS             41602. .42529
                     /locus_tag="DAQB-36F16.1-001"
                     /note="match: proteins: O43626 Q8NGQ2 Q8NGZ5 Q8VEW1
                     Q9GKF5"
                     /pseudo
                     /codon_start=1
                     /product="olfactory receptor, family 2, subfamily H,
                     member 5 pseudogene"
     gene            50201. .50689
                     /locus_tag="DAQB-36F16.3-001"
                     /pseudo
     CDS             50201. .50689
                     /locus_tag="DAQB-36F16.3-001"
                     /note="match: proteins: BAB25132 P35427 Q962U0"
                     /pseudo
                     /codon_start=1
                     /product="ribosomal protein L13a pseudogene"
     gene            55148. .56591
                     /locus_tag="DAQB-36F16.7-001"
```

```
     mRNA            55148. .56591
                     /locus_tag="DAQB-36F16.7-001"
                     /product="olfactory receptor, family 2, subfamily H,
                     member 2"
                     /note="match: cDNAs: AJ459832.1 AJ459839.1 BC036097.1
                     BC048991.1"
     gene            join(55363. .55618,55859. .>56556)
                     /locus_tag="DAQB-36F16.7-002"
     mRNA            join(55363. .55618,55859. .>56556)
                     /locus_tag="DAQB-36F16.7-002"
                     /product="olfactory receptor, family 2, subfamily H,
                     member 2"
                     /note="match: cDNAs: AJ459846.1"
     CDS             55638. .56576
                     /locus_tag="DAQB-36F16.7-001"
                     /standard_name="OTTHUMP00000014633"
                     /note="match: proteins: Q8VF00 Q8VGT6 Q9GZK4"
                     /codon_start=1
                     /product="olfactory receptor, family 2, subfamily H,
                     member 2"
                     /protein_id="CAI18602.1"
                     /db_xref="GI:56207880"
                     /db_xref="GOA:Q5SUK1"
                     /db_xref="InterPro:IPR000276"
                     /db_xref="InterPro:IPR000725"
                     /db_xref="InterPro:IPR002197"
                     /db_xref="UniProt/TrEMBL:Q5SUK1"
                     /translation="MVNQSSTPGFLLLGFSEHPGLERTLFVVVLTSYLLTLVGNTLII
                     LLSALDPKLHSPMYFFLSNLSFLDLCFTTSCVPQMLVNLWGPKKTISFLDCSVQIFIF
                     LSLGTTECILLTVMAFDRYVAVCQPLHYATIIHPRLCWQLASVAWVIGLVESVVQTPS
                     TLHLPFCPDRQVDDFVCEVPALIRLSCEDTSYNEIQVAVASVFILVVPLSLILVSYGA
                     ITWAVLRINSAKGRRKAFGTCSSHLTVVTLFYSSVIAVYLQPKNPYAQERGKFFGLFY
                     AVGTPSLNPLIYTLRNKEVTRAFRRLLGKEMGLTQS"
     CDS             55872. .>56556
                     /locus_tag="DAQB-36F16.7-002"
                     /standard_name="OTTHUMP00000014634"
                     /codon_start=1
                     /product="olfactory receptor, family 2, subfamily H,
                     member 2"
                     /protein_id="CAI18603.1"
                     /db_xref="GI:56207881"
                     /db_xref="GOA:Q5RJ87"
                     /db_xref="InterPro:IPR000276"
                     /db_xref="InterPro:IPR000725"
                     /db_xref="InterPro:IPR002197"
                     /db_xref="UniProt/TrEMBL:Q5RJ87"
                     /translation="MLVNLWGPKKTISFLDCSVQIFIFLSLGTTECILLTVMAFDRYV
                     AVCQPLHYATIIHPRLCWQLASVAWVIGLVESVVQTPSTLHLPFCPDRQVDDFVCEVP
                     ALIRLSCEDTSYNEIQVAVASVFILVVPLSLILVSYGAITWAVLRINSAKGRRKAFGT
                     CSSHLTVVTLFYSSVIAVYLQPKNPYAQERGKFFGLFYAVGTPSLNPLIYTLRNKEVT
                     RAFRRLLGKE"
     gene            complement(join(69921. .71409,72187. .72330,72553. .72681,
                     73262. .73389,74084. .74177,74590. .74697,74795. .74911,
                     76294. .76426,76922. .77072,78617. .78694,80244. .80307,
                     80936. .81178,88795. .88986,89446. .89511,89798. .89899,
                     90999. .91169,91584. .91718,95180. .95340,96781. .96801,
                     98152. .98337,99090. .99293,99995. .100079,100544. .100854))
                     /gene="GABBR1"
                     /locus_tag="DAQB-36F16.2-001"
     mRNA            complement(join(69921. .71409,72187. .72330,72553. .72681,
                     73262. .73389,74084. .74177,74590. .74697,74795. .74911,
                     76294. .76426,76922. .77072,78617. .78694,80244. .80307,
                     80936. .81178,88795. .88986,89446. .89511,89798. .89899,
                     90999. .91169,91584. .91718,95180. .95340,96781. .96801,
                     98152. .98337,99090. .99293,99995. .100079,100544. .100854))
                     /gene="GABBR1"

Alignment Scores:
Pred. No.:              2.02e+04        Length:         102684
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0

DB:                     8               Gaps:           0

US-09-858-332G-28 (1-6) x BX000688 (1-102684)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      58968 CATAATCATAATCATAAT 58985

RESULT 571
AP008086/c
LOCUS       AP008086              103011 bp    DNA     linear   HTG 28-DEC-2004
DEFINITION  Lotus corniculatus var. japonicus chromosome 1 clone LjT21D14, ***
            SEQUENCING IN PROGRESS ***, 25 unordered pieces.
ACCESSION   AP008086
VERSION     AP008086.1  GI:56806393
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1
  AUTHORS   Kaneko,T., Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Structural Analysis of a Lotus japonicus Genome. XI. Sequence
            Features and Mapping of Nine hundred twenty-one TAC Clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 103011)
  AUTHORS   Sato,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-2004) Shusei Sato, Kazusa DNA Research Institute,
            Department of Plant Gene Research; 2-6-7 Kazusa-kamatari, Kisarazu,
            Chiba, 292-0818, Japan (E-mail:ssato@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/, Tel:81-438-52-3935(ex.2337),
            Fax:81-438-52-3934)
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 25 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      337: contig of 337 bp in length
            *      338      437: gap of unknown length
            *      438     1608: contig of 1171 bp in length
            *     1609     1708: gap of unknown length
            *     1709     3128: contig of 1420 bp in length
            *     3129     3228: gap of unknown length
            *     3229     4032: contig of 804 bp in length
            *     4033     4132: gap of unknown length
            *     4133     5629: contig of 1497 bp in length
            *     5630     5729: gap of unknown length
            *     5730     7114: contig of 1385 bp in length
            *     7115     7214: gap of unknown length
            *     7215     8733: contig of 1519 bp in length
            *     8734     8833: gap of unknown length
            *     8834     9894: contig of 1061 bp in length
            *     9895     9994: gap of unknown length
            *     9995    11784: contig of 1790 bp in length
            *    11785    11884: gap of unknown length
            *    11885    13433: contig of 1549 bp in length
            *    13434    13533: gap of unknown length
            *    13534    16020: contig of 2487 bp in length
            *    16021    16120: gap of unknown length
            *    16121    19681: contig of 3561 bp in length
            *    19682    19781: gap of unknown length
            *    19782    24225: contig of 4444 bp in length
            *    24226    24325: gap of unknown length
            *    24326    28200: contig of 3875 bp in length
            *    28201    28300: gap of unknown length
            *    28301    32754: contig of 4454 bp in length
            *    32755    32854: gap of unknown length
```

```
            *    32855    37439: contig of 4585 bp in length
            *    37440    37539: gap of unknown length
            *    37540    42391: contig of 4852 bp in length
            *    42392    42491: gap of unknown length
            *    42492    47248: contig of 4757 bp in length
            *    47249    47348: gap of unknown length
            *    47349    55004: contig of 7656 bp in length
            *    55005    55104: gap of unknown length
            *    55105    61326: contig of 6222 bp in length
            *    61327    61426: gap of unknown length
            *    61427    68710: contig of 7284 bp in length
            *    68711    68810: gap of unknown length
            *    68811    75544: contig of 6734 bp in length
            *    75545    75644: gap of unknown length
            *    75645    83528: contig of 7884 bp in length
            *    83529    83628: gap of unknown length
            *    83629    93084: contig of 9456 bp in length
            *    93085    93184: gap of unknown length
            *    93185   103011: contig of 9827 bp in length.
FEATURES             Location/Qualifiers
     source          1. .103011
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="genomic DNA"
                     /variety="japonicus"
                     /db_xref="taxon:34305"
                     /chromosome="1"
                     /clone="LjT21D14"
                     /clone_lib="LjT library"
                     /note="TAC clone:TM1447, synonym:Lotus japonicus"
     gap             338. .437
                     /estimated_length=unknown
     gap             1609. .1708
                     /estimated_length=unknown
     gap             3129. .3228
                     /estimated_length=unknown
     gap             4033. .4132
                     /estimated_length=unknown
     gap             5630. .5729
                     /estimated_length=unknown
     gap             7115. .7214
                     /estimated_length=unknown
     gap             8734. .8833
                     /estimated_length=unknown
     gap             9895. .9994
                     /estimated_length=unknown
     gap             11785. .11884
                     /estimated_length=unknown
     gap             13434. .13533
                     /estimated_length=unknown
     gap             16021. .16120
                     /estimated_length=unknown
     gap             19682. .19781
                     /estimated_length=unknown
     gap             24226. .24325
                     /estimated_length=unknown
     gap             28201. .28300
                     /estimated_length=unknown
     gap             32755. .32854
                     /estimated_length=unknown
     gap             37440. .37539
                     /estimated_length=unknown
     gap             42392. .42491
                     /estimated_length=unknown
     gap             47249. .47348
                     /estimated_length=unknown
     gap             55005. .55104
                     /estimated_length=unknown
     gap             61327. .61426
                     /estimated_length=unknown
     gap             68711. .68810
                     /estimated_length=unknown
     gap             75545. .75644
                     /estimated_length=unknown
     gap             83529. .83628
                     /estimated_length=unknown
     gap             93085. .93184
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:                2.03e+04        Length:          103011
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       14              Gaps:            0

US-09-858-332G-28 (1-6) x AP008086 (1-103011)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      27372 CATAATCATAATCATAAT 27355

RESULT 572
CR352288
LOCUS       CR352288              103979 bp    DNA     linear   VRT 19-MAR-2005
DEFINITION  Zebrafish DNA sequence from clone CH211-123P18 in linkage group 16,
            complete sequence.
ACCESSION   CR352288
VERSION     CR352288.9  GI:61657063
KEYWORDS    HTG.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 103979)
  AUTHORS   Gray,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Mar 19, 2005 this sequence version replaced gi:58651979.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived
            Zebrafish pUC subclones occasionally display inconsistency over the
            length of mononucleotide A/T runs and conserved TA repeats. Where
            this is found the longest good quality representation will be
            submitted.
            Repeat names beginning 'Dr' were identified by the Recon repeat
            discovery system (Zhirong Bao and Sean Eddy, submitted), and those
            beginning 'drr' were identified by Rick Waterman (Stephen Johnson
            lab, WashU).  For further information see
```

```
           http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml
           CH211-123P18 is from a CHORI-211 BAC library
           VECTOR: pTARBAC2.1.
FEATURES             Location/Qualifiers
     source          1. .103979
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-123P18"
                     /clone_lib="CHORI-211"
ORIGIN

Alignment Scores:
Pred. No.:              2.05e+04        Length:         103979
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     5               Gaps:           0

US-09-858-332G-28 (1-6) x CR352288 (1-103979)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      63245 CATAATCATAATCATAAT 63262

RESULT 573
AP008156
LOCUS       AP008156               104423 bp    DNA     linear   HTG 28-DEC-2004
DEFINITION  Lotus corniculatus var. japonicus chromosome 1 clone LjT17N12, ***
            SEQUENCING IN PROGRESS ***, 22 unordered pieces.
ACCESSION   AP008156
VERSION     AP008156.1  GI:56806462
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1
  AUTHORS   Kaneko,T., Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Structural Analysis of a Lotus japonicus Genome. XI. Sequence
            Features and Mapping of Nine hundred twenty-one TAC Clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 104423)
  AUTHORS   Sato,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-2004) Shusei Sato, Kazusa DNA Research Institute,
            Department of Plant Gene Research; 2-6-7 Kazusa-kamatari, Kisarazu,
            Chiba, 292-0818, Japan (E-mail:ssato@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/, Tel:81-438-52-3935(ex.2337),
            Fax:81-438-52-3934)
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 22 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      153: contig of 153 bp in length
            *      154      253: gap of unknown length
            *      254      994: contig of 741 bp in length
            *      995     1094: gap of unknown length
            *     1095     2544: contig of 1450 bp in length
            *     2545     2644: gap of unknown length
            *     2645     4746: contig of 2102 bp in length
            *     4747     4846: gap of unknown length
            *     4847     7076: contig of 2230 bp in length
            *     7077     7176: gap of unknown length
            *     7177     8526: contig of 1350 bp in length
            *     8527     8626: gap of unknown length
            *     8627    10602: contig of 1976 bp in length
            *    10603    10702: gap of unknown length
            *    10703    12246: contig of 1544 bp in length
            *    12247    12346: gap of unknown length
            *    12347    14006: contig of 1660 bp in length
            *    14007    14106: gap of unknown length
            *    14107    16013: contig of 1907 bp in length
            *    16014    16113: gap of unknown length
            *    16114    17651: contig of 1538 bp in length
            *    17652    17751: gap of unknown length
            *    17752    21610: contig of 3859 bp in length
            *    21611    21710: gap of unknown length
            *    21711    25518: contig of 3808 bp in length
            *    25519    25618: gap of unknown length
            *    25619    29775: contig of 4157 bp in length
            *    29776    29875: gap of unknown length
            *    29876    35483: contig of 5608 bp in length
            *    35484    35583: gap of unknown length
            *    35584    41669: contig of 6086 bp in length
            *    41670    41769: gap of unknown length
            *    41770    48039: contig of 6270 bp in length
            *    48040    48139: gap of unknown length
            *    48140    55825: contig of 7686 bp in length
            *    55826    55925: gap of unknown length
            *    55926    66035: contig of 10110 bp in length
            *    66036    66135: gap of unknown length
            *    66136    76587: contig of 10452 bp in length
            *    76588    76687: gap of unknown length
            *    76688    88811: contig of 12124 bp in length
            *    88812    88911: gap of unknown length
            *    88912   104423: contig of 15512 bp in length.
FEATURES             Location/Qualifiers
     source          1. .104423
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="genomic DNA"
                     /variety="japonicus"
                     /db_xref="taxon:34305"
                     /chromosome="1"
                     /clone="LjT17N12"
                     /clone_lib="LjT library"
                     /note="TAC clone:TM1564, synonym:Lotus japonicus"
     gap             154. .253
                     /estimated_length=unknown
     gap             995. .1094
                     /estimated_length=unknown
     gap             2545. .2644
                     /estimated_length=unknown
     gap             4747. .4846
                     /estimated_length=unknown
     gap             7077. .7176
                     /estimated_length=unknown
     gap             8527. .8626
                     /estimated_length=unknown
     gap             10603. .10702
                     /estimated_length=unknown
     gap             12247. .12346
                     /estimated_length=unknown
     gap             14007. .14106
                     /estimated_length=unknown
     gap             16014. .16113
                     /estimated_length=unknown
     gap             17652. .17751
                     /estimated_length=unknown
     gap             21611. .21710
                     /estimated_length=unknown
     gap             25519. .25618
                     /estimated_length=unknown
     gap             29776. .29875
                     /estimated_length=unknown
     gap             35484. .35583
                     /estimated_length=unknown
     gap             41670. .41769
                     /estimated_length=unknown
```

```
     gap             48040. .48139
                     /estimated_length=unknown
     gap             55826. .55925
                     /estimated_length=unknown
     gap             66036. .66135
                     /estimated_length=unknown
     gap             76588. .76687
                     /estimated_length=unknown
     gap             88812. .88911
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              2.06e+04        Length:         104423
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AP008156 (1-104423)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      39787 CACAATCACAATCACAAT 39804

RESULT 574
BX469894
LOCUS       BX469894              104775 bp    DNA     linear   VRT 27-JUL-2004
DEFINITION  Zebrafish DNA sequence from clone CH211-212N23 in linkage group 14,
            complete sequence.
ACCESSION   BX469894
VERSION     BX469894.4  GI:50724588
KEYWORDS    HTG.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 104775)
  AUTHORS   Leongamornlert,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 27, 2004 this sequence version replaced gi:32959713.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived
            Zebrafish pUC subclones occasionally display inconsistency over the
            length of mononucleotide A/T runs and conserved TA repeats. Where
            this is found the longest good quality representation will be
            submitted.
            Repeat names beginning 'Dr' were identified by the Recon repeat
            discovery system (Zhirong Bao and Sean Eddy, submitted), and those
            beginning 'drr' were identified by Rick Waterman (Stephen Johnson
            lab, WashU).  For further information see
            http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml
            CH211-212N23 is from a CHORI-211 BAC library
            VECTOR: pTARBAC2.1.
FEATURES                Location/Qualifiers
     source             1. .104775
                        /organism="Danio rerio"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:7955"
                        /clone="CH211-212N23"
                        /clone_lib="CHORI-211"
ORIGIN

Alignment Scores:
Pred. No.:              2.06e+04        Length:         104775
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     5               Gaps:           0

US-09-858-332G-28 (1-6) x BX469894 (1-104775)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      90823 CATAACCACAACCACAAT 90840

RESULT 575
AC137065/c
LOCUS       AC137065              105004 bp    DNA     linear   PLN 08-APR-2004
DEFINITION  Medicago truncatula clone mth2-29h7, complete sequence.
ACCESSION   AC137065
VERSION     AC137065.26  GI:46275904
KEYWORDS    HTG.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 105004)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Medicago truncatula BAC Clone mth2-29h7
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 105004)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 105004)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-2004) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   4  (bases 1 to 105004)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-2004) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Apr 8, 2004 this sequence version replaced gi:45774224.
            -------------- Genome Center
```

```
           Center: Department Of Chemistry And Biochemistry
           The University Of Oklahoma
           Center code:UOKNOR
           ----------------------.
FEATURES             Location/Qualifiers
     source          1. .105004
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3880"
                     /clone="mth2-29h7"
                     /clone_lib="Medicago truncatula BAC library H2"
ORIGIN

Alignment Scores:
Pred. No.:              2.07e+04       Length:         105004
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     15             Gaps:           0

US-09-858-332G-28 (1-6) x AC137065 (1-105004)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      64014 CATAATCACAACCATAAC 63997

RESULT 576
AL390959/c
LOCUS       AL390959              105239 bp    DNA     linear   PRI 18-MAY-2005
DEFINITION  Human DNA sequence from clone RP1-44L3 on chromosome 6 Contains a
            CpG island, complete sequence.
ACCESSION   AL390959
VERSION     AL390959.12  GI:11967559
KEYWORDS    HTG; CpG island.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 105239)
  AUTHORS   Tracey,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk
            Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Dec 22, 2000 this sequence version replaced gi:11611046.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP1-44L3 is from the library RPCI-1 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: vega@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one subclone; and the assembly was confirmed by restriction digest,
            except on the rare occasion of the clone being a YAC.

FEATURES             Location/Qualifiers
     source          1. .105239
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="RZPD:RPCIP704L0344"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP1-44L3"
                     /clone_lib="RPCI-1"
     misc_feature    100
                     /note="Clone_right_end: RP4-626B19"
     misc_feature    105239
                     /note="Clone_right_end: RP1-44L3"
ORIGIN

Alignment Scores:
Pred. No.:              2.07e+04       Length:         105239
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     8              Gaps:           0

US-09-858-332G-28 (1-6) x AL390959 (1-105239)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      19535 CATAATCATAATCATAAT 19518

RESULT 577
AC126792/c
LOCUS       AC126792              105297 bp    DNA     linear   PLN 22-MAR-2005
DEFINITION  Medicago truncatula clone mth2-8o8, complete sequence.
ACCESSION   AC126792
VERSION     AC126792.15  GI:61675754
KEYWORDS    HTG.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 105297)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Medicago truncatula BAC Clone mth2-8o8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 105297)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 105297)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   4  (bases 1 to 105297)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-SEP-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   5  (bases 1 to 105297)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2002) Department Of Chemistry And Biochemistry,
```

```
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   6  (bases 1 to 105297)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   7  (bases 1 to 105297)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-APR-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   8  (bases 1 to 105297)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   9  (bases 1 to 105297)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   10 (bases 1 to 105297)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAY-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   11 (bases 1 to 105297)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUN-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   12 (bases 1 to 105297)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-OCT-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   13 (bases 1 to 105297)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JAN-2004) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   14 (bases 1 to 105297)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUN-2004) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   15 (bases 1 to 105297)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-2005) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Mar 22, 2005 this sequence version replaced gi:48717536.
            ---------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ----------.
FEATURES             Location/Qualifiers
     source          1. .105297
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3880"
                     /clone="mth2-8o8"
                     /clone_lib="Medicago truncatula BAC library H1"
ORIGIN

Alignment Scores:
Pred. No.:                2.07e+04        Length:          105297
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       15              Gaps:            0

US-09-858-332G-28 (1-6) x AC126792 (1-105297)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      99166 CACAATCACAATCACAAT 99149

RESULT 578
AC116920
LOCUS       AC116920               105320 bp    DNA     linear   INV 21-FEB-2004
DEFINITION  Dictyostelium discoideum chromosome 2 map 4415041-4520361 strain
            AX4, complete sequence.
ACCESSION   AC116920 AC116983
VERSION     AC116920.3  GI:42733646
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 105320)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 105320)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 105320)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 105320)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   5  (bases 1 to 105320)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2004) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On Feb 21, 2004 this sequence version replaced gi:28828137.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
```

```
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES    Location/Qualifiers
     source 1. .105320
            /organism="Dictyostelium discoideum"
            /mol_type="genomic DNA"
            /strain="AX4"
            /db_xref="taxon:44689"
            /chromosome="2"
            /map="4415041-4520361"
     CDS    join(<84. .195,401. .445)
            /note="GeneID exon scores (in order of location ranges):
            5.39, -2.56 - GSCJ_ID dd_03376"
            /codon_start=2
            /product="hypothetical protein"
            /protein_id="AAS38624.1"
            /db_xref="GI:42733679"
            /translation="IEILASSTKHFQSICRRLYRIFAHAFYHHREAYNEYENSLLPKS
            AIIIKEN"
     CDS    join(2017. .2175,2265. .2273)
            /note="GeneID exon scores (in order of location ranges):
            16.22, -1.64 - GSCJ_ID dd_00543"
            /codon_start=1
            /product="similar to Homology to rat S29; Rps29bp
            [Saccharomyces cerevisiae]"
            /protein_id="AAS38610.1"
            /db_xref="GI:42733647"
            /translation="MARELWLTHPRNFGPGSRTCRKCGNHHGIIRKYDLNMCRRCFRT
            DAEAIGFNKYR"
     CDS    complement(2879. .3808)
            /note="GeneID exon scores (in order of location ranges):
            24.99 - GSCJ_ID dd_00545"
            /codon_start=1
            /product="hypothetical protein"
            /protein_id="AAS38611.1"
            /db_xref="GI:42733648"
            /translation="MVEQCPLDMCLSYNLCTSNSSFGNLYFGCKDEYGNKVQTKNGIR
            CFSHFKPILLIPILVFHLIFALVLSLIKHNSFRIPISQIIIFLQINSVVFFSYSGIYI
            LPLFGMSVDMLDGYCLWEGLNYNFKSFLSFSIIISLFLIGSTDITPIVLIKIYEKIPL
            IANNNSIRNRKKNINDSEKTIIEKTWCPRFLKEKIKKRINEKSNSKVYYWKSIWSFIS
            TFCNTFDVQLNIFNTILYCYYTSIQIVYMSFYFFLATSSNNSSRDLNNYINLAQLIIF
            IVANSNLLQLGSLFGQGLVLTSVSFSFSLVLIN"
     CDS    complement(join(4065. .4273,4441. .4587,5023. .5188))
            /note="GeneID exon scores (in order of location ranges):
            8.11, 1.77, -2.21 - GSCJ_ID dd_00546"
            /codon_start=1
            /product="hypothetical protein"
            /protein_id="AAS38612.1"
            /db_xref="GI:42733649"
            /translation="MIISENASLIIFNSIKLNENVGSFLKSSNDKSLHINTLIYNDNK
            IPYDYLFSIENYLIASFNIAKFGGVFQFGGAYSNKNDAFINPISKFINNKFSSNIGSQ
            QGELQKAIFSNDSINVINGVDTLVKFYFFDYFNNTFASLSMPLNISITRLLDSSTKFS
            LSIMAGKGIPLSV"
     CDS    complement(join(5249. .6306,6524. .6773))
            /note="GeneID exon scores (in order of location ranges):
            67.43, 19.81 - GSCJ_ID dd_00548"
            /codon_start=1
            /product="hypothetical protein"
            /protein_id="AAS38613.1"
            /db_xref="GI:42733650"
            /translation="MVFVNSTFRFTVFTGELNATNSLFRFIGSSFINGGIVFNLINSD
            SSSGDSSSNSSSNSSSNSSSSSNSNIICISGCGVSFSDNNTTITFSKSVDPLIIIIAY
            RYSYVSIENVKIKGYRSSNPMVISSEGSEIIMNSIFIEDCISNSLISTRKGGFSDGGS
            IEINKIKFKDNTFDSIIDVTGVSNVTISELITTCEMNNKFYSMIVMATADDVYDVNNQ
            YLSKSNVSFSFMELSCIGSFQFERALVELLIVNMSASFNQFLFTLTKESILVIDRCDF
            ILDPLTCSTKSTAFLVENISDLYIVDSIFGYSGTGIVVSTNGSSITLVNTRILSIILS
            VPVFTIGHSTLDFYNCTFEQTYIPSTSLIDSFISNITIDICTIDTLGDVFLDGFGSNI
            IINQLVMTNFFTQYRLFQSNSDSIIISNSLFINIHGFGYQLFQ"
     CDS    complement(join(6894. .7206,7397. .7407))
            /note="GeneID exon scores (in order of location ranges):
            4.13, -2.65 - GSCJ_ID dd_00551"
            /codon_start=1
            /product="hypothetical protein"
            /protein_id="AAS38614.1"
            /db_xref="GI:42733651"
            /translation="MKKFNNESNGTIINPFQDLCILNNILQQSFYNFNNNYYENINFN
            QNNLNSIIKIYIGYGNYNLTCQSISISNFNNISISPYNFNQNQTIIENLFLNLENTNL
            IINGP"
     CDS    complement(join(7649. .9423,9499. .12796))
            /note="GeneID exon scores (in order of location ranges):
            89.47, 176.06 - GSCJ_ID dd_00554"
            /codon_start=1
            /product="similar to Plasmodium falciparum. Hypothetical
            protein"
            /protein_id="AAS38615.1"
            /db_xref="GI:42733652"
            /translation="MKIYKYSLTFIFLIFIFLNNVCSQNINNNNNNINQYEDNNIINI
            YINSNNQNNGNGTIINPYNSICQIESILNNINNNNNNNNNKNNNNIATSIIIVNIGY
            GIYNDTNCYQFLILNCKNVTFKPYISNSSSSKDNNNNNNNNNNNNNNNNNNNNNNNN
            KEEEEGEQKVSIYNSFKIINSKFYIQGPFITFYGSWNISNSEMNIIDNAYIIFKDKYN
            IYSIISNSVIKLINSTMISSIYDQKYMIEDSLLYFQKSTLKQVSINFNSNPKFASSGG
            GGRVEIIDSKVMGSFFQMGFSDINVKNSIWQETVLFTVADNSKFENILFNQNDLKLWS
            EYYNSVGIINIDKELFNENGSKITILNSEFRANPKYGALSLIVSLDIEQDVLIILYNI
            RMIGFHKYESMFASTSYGGNITISLELVEFDNCVSPSMFTMENTNLVISDLIFTNCIF
            ETMLEARNSSNIEISNLSINSTQNNIIQKFLSMPKDENSFEMSSSIEFSNIFLESIEG
            SFEFRNSIVNFEFVTMISIGSDSFINGYQGSIVNIISSDFISNPFGKINQLFYFEQSS
            ITISQSNFFDSAVPIVMAVKNSSVLIDNNSKFIDLQTSQPIITVADDSILEIYSTLFN
            NCRISSSSLIQSESSLVYISETNFLLVSQQLFNSSRSKISMNIVNVSQSFTQSQLITS
            VDDYIFIASSVFDGNHGLSYQMFTLNGSILNITQTQFNGNSFNPLIKSVNSTIEFVET
            LYSNNTGSFINSKDDQLISISFLEFYDSRLPINYLLSSDSTKTIILNQIFINNNYFTQ
            FFISSFNAETFEIDTFEFGNNLIVHVPMLLCPYNQFDGLGGFISLTQVSNVLIKNFNI
            RNNQIGTTFLTFYDCVFSKYNSNNSNSNSYSNSYSNTGSGVGGGSSGSCSGSSSGSSS
            GSSSCSGSVNIQNSQFTSNTIQSDISLYSESNYLIESKSSEVSISLTYFNGNQGNGKG
            LILVINSNITIYGCSIDSNFFENGTGAILSFFPDYLIDDSFQWFVIVQSLLTTNFGIY
            GGVLYFNSPQSLNNPLSKIENNSFINNRALQQGGVVYTTDSPFNENMTKVLNSNFFIG
            NTAFCGPNFSSNMMKNPQLSNPISQVVDDGSSIIEYQFVDYYGNLFTPFSSILTIYVN
            RTSTNHLMNVTNNVNAGMGKFTFEFIGSINDEYLIGVNSPTPVLAKVYVDGCSPYQFY
            NTTSEMCTYCPNNYALSQLLGKCYPCNTNIMNCLASTVFAIDGYYLVNNDVSMVEQCP
            LDMCLSYNLCTSNSSFGNLCFSCKDEYGNLVQTKNGIRCCSHFKPILLIPILVFHLIF
            ALVLSLIKHNSFRIPISQIIIFLQINSVVFFSYSGIYILPLFRMSVDMLDGYCLWEGL
            NYNFKSFLSFSIIISLFLIGSTDITPSIIIKIYKKISTRKDNRNLSLLINDRNITPLQ
            QQQQPIAKYPNYIIKILEGKIKHRSIYYWKSIWSLFQLFVIPLAFNAMSLVISKQVGS
            ENYLAVDFSIDFQSNSNQACILIAILIFIILAIVLFMQLILLLSKFKIGGNRIFKNRI
            TESLLKIQQYQLEYRTNFKWFPIVINSRSIIICVLSIVLLFKAQYYIPIILLIQIFYT
            IISLLINPIKWKYLRDFNNIINLLQVVILITANSNLLQSKSLFNEGLIITIITFSFAL
            LSIISIFLLYNGKNKKIRRKLK"
     CDS    join(13910. .13922,14046. .14224)
            /note="GeneID exon scores (in order of location ranges):
            2.36, 2.61 - GSCJ_ID dd_00558"
            /codon_start=1
            /product="hypothetical protein"
            /protein_id="AAS38616.1"
            /db_xref="GI:42733653"
            /translation="MPILGSISTFSQFNVKKSKNDSISSFLKSNVGYSNSMVSEIDGR
            FDNLPIIIQKPPNKPNAYI"
     CDS    complement(join(14483. .19185,19254. .19814,19913. .19922))
            /note="GeneID exon scores (in order of location ranges):
            556.18, 66.59, 0.22 - GSCJ_ID dd_00562"
            /codon_start=1
            /product="similar to unknown protein; protein id:
            At1g65440.1 [Arabidopsis thaliana]"
            /protein_id="AAO50823.2"
            /db_xref="GI:42733654"
            /translation="MSSKNYSDDEIENEENIHPKKKFKNLNKNKDDESSELEEYHDSS
            EEDENEPNQYQYDDFVVGGDDDERRKRKKKEKRKKSSSNSDNEDDNEEEEEEDRGGGG
            GSGSGGSGGSGGRNSSSPKRTNNHLDEDIDGDEDGDDGSDGSESESESEEDEEVAQAR
            RERRRKRREERRKREKGGLQRLKKLRASNKDLSDKEDDVEDGYNGEQEDDGFVVDSHG
            KPVKRSHERERMSYESSNRSRIMSDIFDDDEEEGYGSGEEEEDSGINKTEADRAKDRL
            AAIREQYEPSLLEEKHFTDADEEIRNKNVPERLQSRKGTQYAGEMACEEEAEWIYEVA
            FEQRDFQQQQQRNPNTEPVRDAKAITAIFKILQFIQRDLLEIPFIYTYEKDIYEPYFT
            LQELWNIFDLDEKWAHMKVNKRNLEQIGSNNQQLEPYKSVLLESRSEESISDLYDLFQ
            MINGIQKNTNLLNGSSSLGSGGDGYDDKEGGSSATQGNTVGGLSRQKKAIKRDLYTIY
            TKAGLSKFLPNFGMTAQEFGQNLMDNYTTNKPRDQMAEPSTMALVHICLEADNKDRVL
            QATRYMMAQEIGYDPHVRYSVRMIYRKYAHITTTPTLKGFKEIDVFHPYFTVKSIQEK
```

```
                     PAHLFEDSQYLLILKAEKEGFIKSTMAITEKTHNSVIIPEMESLYLSDGTSAITQQWN
                     EQRKLIIREALNKFLYPVLEKELRNKLLTEASNRVAFECAKKLEDKIRVAPWRPVSSS
                     SGIGHSNSSVLFGGSRANNGYSNDDDEDEDDDEEVDYDRDGNPIPFKILSLCWGSDKI
                     PTMGAVLNSDGEVLSHIKLDFICDRLGESLKEKKEKDIKRLEEICQEYQPRLVLVSAT
                     EMDSKRLYEEVKDHLQRWSNGERRIIRKSVLLNYYSPEIGLSLQTSSRLQEEFKEYPP
                     ILRHTIAVGRCALDPITEYASLCNDHNEILFLKLHPLQDMIGKDYLVKLLHRCFINVV
                     NAVGVDINRMIQCRFTSATLQFVSGLGSRKAQMLLNSLFRRGGFVTSRQSLEKVLSQD
                     VIYRNCIGFIQIRERHAADYKADLLDDTRIHPTSYPITYRIAAEALDKNLDERNFHSY
                     IEDIIKRPKKLDRLDLDGFADIIESHEDRPARKLLYFIKKELTNPFADIRHPYEEPSA
                     DTIFEWLTGETNQSLRRGTLVTVTTIRTFDNSVKCRLDNGLEGSIPTDCISDNGETKS
                     LGRGVTINCRVMGVDKYQFTVSLSCKPSDLNPSFWEETIFRELKENGANQYLRLEEVA

Alignment Scores:
Pred. No.:                2.07e+04        Length:         105320
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       2               Gaps:           0

US-09-858-332G-28 (1-6) x AC116920 (1-105320)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      36960 CACAATCACAATCACAAC 36977

RESULT 579
AC150706/c
LOCUS       AC150706               105472 bp    DNA     linear   PLN 06-NOV-2004
DEFINITION  Medicago truncatula clone mth2-132l18, complete sequence.
ACCESSION   AC150706
VERSION     AC150706.4  GI:55468707
KEYWORDS    HTG.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 105472)
  AUTHORS   Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
            and Roe,B.A.
  TITLE     Medicago truncatula BAC Clone mth2-132l18
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 105472)
  AUTHORS   Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
            and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-2004) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 105472)
  AUTHORS   Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
            and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2004) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   4  (bases 1 to 105472)
  AUTHORS   Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
            and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-OCT-2004) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   5  (bases 1 to 105472)
  AUTHORS   Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
            and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-2004) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Nov 6, 2004 this sequence version replaced gi:54888778.
            ---------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ----------.
FEATURES             Location/Qualifiers
     source          1. .105472
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3880"
                     /clone="mth2-132l18"
                     /clone_lib="Medicago truncatula BAC library H2"
ORIGIN

Alignment Scores:
Pred. No.:                2.08e+04        Length:         105472
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       15              Gaps:           0

US-09-858-332G-28 (1-6) x AC150706 (1-105472)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      44450 CATAACCATAACCATAAT 44433

RESULT 580
AC148819/c
LOCUS       AC148819               105682 bp    DNA     linear   PLN 13-MAY-2004
DEFINITION  Medicago truncatula chromosome 2 BAC clone mte1-11g7, complete
            sequence.
ACCESSION   AC148819
VERSION     AC148819.3  GI:47155624
KEYWORDS    HTG.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 105682)
  AUTHORS   Town,C.D., Tallon,L.J., Arbogast,T., Althoff,R., Hine,E.,
            Monaghan,E., Smith,S.A., Utterback,T., Feldblyum,T., Koo,H. and
            Cheung,F.
  TITLE     Medicago truncatula BAC genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 105682)
  AUTHORS   Town,C.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-APR-2004) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 105682)
  AUTHORS   Town,C.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAY-2004) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   4  (bases 1 to 105682)
  AUTHORS   Town,C.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2004) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
COMMENT     On May 13, 2004 this sequence version replaced gi:47077941.
FEATURES             Location/Qualifiers
     source          1. .105682
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3880"
                     /chromosome="2"
                     /clone="mte1-11g7"
ORIGIN
```

```
Alignment Scores:
Pred. No.:              2.08e+04        Length:         105682
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x AC148819 (1-105682)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      12401 CATAATCATAATCATAAT 12384

RESULT 581
AC069474/c
LOCUS       AC069474              105768 bp    DNA     linear   PLN 19-JAN-2001
DEFINITION  Arabidopsis thaliana chromosome 3 BAC T2E22 genomic sequence,
            complete sequence.
ACCESSION   AC069474
VERSION     AC069474.4  GI:12321944
KEYWORDS    HTG.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 105768)
  AUTHORS   Lin,X., Kaul,S., Town,C.D., Benito,M.-I., Creasy,T.H., Haas,B.J.,
            Wu,D., Maiti,R., Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R.,
            Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.
  TITLE     Arabidopsis thaliana chromosome 3 BAC T2E22 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 105768)
  AUTHORS   Town,C.D. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUN-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
REFERENCE   3  (bases 1 to 105768)
  AUTHORS   Town,C.D. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
COMMENT     On Jan 19, 2001 this sequence version replaced gi:12280877.
            Address all correspondence to:at@tigr.org

            BAC clone T2E22 is from Arabidopsis thaliana chromosome 3
            The orientation of the sequence is from SP6 to T7 end of the BAC
            clone.
            Genes were identified by a combination of several methods: Gene
            prediction programs including Genscan+ (Chris Burge,
            http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHMM (Mark Borodovsky,
            http://genemark.biology.gatech.edu/GeneMark/), GlimmerA (a variant
            of GlimmerM, see Mihaela Pertea,
            http://www.tigr.org/softlab/glimmerm_htm/glimmerm.html, and
            GeneSplicer (Mihaela Pertea and Steven Salzberg, contact
            mpertea@tigr.org), searches of the complete sequence against a
            peptide database and the plant EST database at TIGR
            (http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to
            indicate the level of evidence for their annotation. Genes with
            similarity to other proteins are named after the database hits.
            Genes without significant peptide similarity but with EST
            similarity are named as  unknown  proteins. Genes without protein
            or EST similarity, that are predicted by more than two gene
            prediction programs over most of their length are annotated as
            hypothetical  proteins. Genes encoding tRNAs are predicted by
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
            Simple repeats are identified by repeatmasker (Arian Smit,
            http://ftp.genome.washington.edu/RM/RepeatMasker.html).
FEATURES             Location/Qualifiers
     source          1. .105768
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /chromosome="III"
                     /map="MS2"
                     /clone="T2E22"
                     /ecotype="Columbia"
     gene            146. .2644
                     /gene="T2E22.1"
                     /note="contains multiple zinc fingers Pfam profile:PF00642
                      Zinc finger C-x8-C-x5-C-x3-H type (and similar)"
     mRNA            join(<146. .197,592. .648,741. .812,969. .1025,1128. .1325,
                     1506. .1543,1641. .1687,1788. .1906,1993. .2046,2142. .2282,
                     2366. .2644)
                     /gene="T2E22.1"
     CDS             join(<146. .197,592. .648,741. .812,969. .1025,1128. .1325,
                     1506. .1543,1641. .1687,1788. .1906,1993. .2046,2142. .2282,
                     2366. .2518)
                     /gene="T2E22.1"
                     /codon_start=2
                     /product="zinc finger protein, putative, 5' partial;
                     146-2518"
                     /protein_id="AAG51026.1"
                     /db_xref="GI:12321969"
                     /translation="FDHPIWVPEGGIPDWKEAPVVPNEEYPERPGEPDCPYYIKTQRC
                     KYGSKCKFNHPREEAAVSVETQDSLPERPSEPMCTFYMKTGKCKFGLSCKFHHPKDIQ
                     LPSSSQDIGSSVGLTSEPDATNNPHVTFTPALYHNSKGLPVRSGEVDCPFYLKTGSCK
                     YGATCRYNHPERTAFIPQAAGVNYSLVSSNTANLNLGLVTPATSFYQTLTQPTLGVIS
                     ATYPQRPGQSECDYYMKTGECKFGERCKFHHPADRLSAMTKQAPQQPNVKLSLAGYPR
                     REGALNCPYYMKTGTCKYGATCKFDHPPPGEVMAKTTSEADAAGATNTDTTQ"
     gene            3708. .7642
                     /gene="T2E22.2"
                     /note="similar to CTP synthetase GB:U49350 GI:1515356 [Mus
                     musculus]"
     mRNA            join(<3708. .3825,3989. .4068,4153. .4239,4315. .4365,
                     4452. .4569,4648. .4704,4804. .4868,4952. .5035,5129. .5179,
                     5299. .5361,5468. .5511,5642. .5711,5803. .5865,6003. .6092,
                     6181. .6219,6293. .6520,6609. .6694,6776. .6824,6912. .6992,
                     7104. .7197,7286. .7642)
                     /gene="T2E22.2"
     CDS             join(3708. .3825,3989. .4068,4153. .4239,4315. .4365,
                     4452. .4569,4648. .4704,4804. .4868,4952. .5035,5129. .5179,
                     5299. .5361,5468. .5511,5642. .5711,5803. .5865,6003. .6092,
                     6181. .6219,6293. .6520,6609. .6694,6776. .6824,6912. .6992,
                     7104. .7197,7286. .7443)
                     /gene="T2E22.2"
                     /codon_start=1
                     /product="CTP-synthetase, putative; 3708-7443"
                     /protein_id="AAG51029.1"
                     /db_xref="GI:12321972"
                     /translation="MKYVLVTGGVVSGLGKGVTASSIGLLLQACGLRVTSIKIDPYLN
                     TDAGTMSPFEHGEVFVLDDGGEVDLDLGNYERFLDSTLTRDNNITTGKIYQSVIDKER
                     KGDYLGRTVQVVPHVTDAIQEWIERVANVPVDGKEGPPDVCVIELGGTIGDIESMPFI
                     EALGQFSYKVGPGNFCLVHVSLVPVLSVVGEQKTKPTQHSVRGLRSLGLTPNILACRS
                     TKALEENVKTKLSQFCHVPEVNIVTLYDVPNIWHVPLLLRDQKAHEAILRELNLSNAI
                     KPDLTEWTARTKIYDTLQDPVRIAMVGKYTGLTDSYLSVLKALLHASVACHKKLVIEW
                     VAASDLEEITAQETPDVHKAAWDLLKGADGILVPGGFGDRGVQGKILATKYARENQVP
                     FLGICLGMQLAVVEFARSILGFHDANSTEFEPETSSPCIIFMPEGSTTHMGGTMRLGS
                     RKTYFQVADCKSAKLYGNAKFVDERHRHRYEVNPDMISEIEKAGLSFVGKDETGRRME
                     IVELPSHPYFVGAQFHPEFKSRPGKPSALFLGLIAAASGCLESVLQTGGKVNIVSKNG
                     VANGSAMGKVHQNGNVYSNGNGLHH"
     gene            complement(7967. .8734)
                     /gene="T2E22.3"
     mRNA            complement(<7967. .>8734)
                     /gene="T2E22.3"
     CDS             complement(7967. .8734)
                     /gene="T2E22.3"
                     /codon_start=1
                     /product="hypothetical protein; 8734-7967"
                     /protein_id="AAG51033.1"
                     /db_xref="GI:12321976"
                     /translation="MSSSLTIFFFFFASTFLYTSSNSFNITNILNEHDDFSNFNQLLS
                     ETQLASTINKRQTITVLVVSNGALSSLSGQPTSVIKKILSLHIVLDYYDQKKLKNLSK
                     KTVLLTTLFQSSGLARGQQGFLNATVMKNGDVAFGSAVPGSSLDAQLQDTVAALPFNI
                     SVLHISSAIMIDVKGDNAPTASPLSPVSSPPRPAESPNDDGQDFDEPPSSAPGAAADE
```

```
                PSENAGSANGVSRNDSQPAFAFTLLMSFIWWFMARLR"
     gene       9197. .9913
                /gene="T2E22.4"
                /note="predicted by genemark.hmm"
     mRNA       <9197. .>9913
                /gene="T2E22.4"
     CDS        9197. .9913
                /gene="T2E22.4"
                /codon_start=1
                /product="hypothetical protein; 9197-9913"
                /protein_id="AAG51037.1"
                /db_xref="GI:12321980"
                /translation="MMISRRLVSKFLKPLSSPSFHVSRRHLQSSSSLIRNHLIHGSVT
                KELGLLRSFSVFNIDKLRSETWSPRYFSTPSGETEPKKPEESKEEELQKMKHQEIEGP
                TVERDLSALGNETRQVLEGMMKNMYSLSGAMGALGLTQLILGAAILYATRADPMKEMT
                IQSCIAFGFPFAMALMVRRSLKPMYFFKKMEELGRLQILTLTLQVAKNLNLLFVRARV
                VSILCVVALCFGNLFLLLSP"
     gene       complement(10119. .13339)
                /gene="T2E22.5"
     mRNA       complement(join(<10119. .10382,10500. .10587,10686. .10792,
                10888. .10949,11103. .11157,11364. .12186,12320. .12357,
                12491. .12842,12924. .13024,13205. .>13339))
                /gene="T2E22.5"
     CDS        complement(join(10119. .10382,10500. .10587,10686. .10792,
                10888. .10949,11103. .11157,11364. .12186,12320. .12357,
                12491. .12842,12924. .13024,13205. .13339))
                /gene="T2E22.5"
                /codon_start=1
                /product="unknown protein; 13339-10119"
                /protein_id="AAG51005.1"
                /db_xref="GI:12321948"
                /translation="MGSADLVDDRTFNGDFSVEGVAKLKEMVKDKLKEYMGDYTDDIL
                VEYVIVLLRNGRRKEEANNELKIFLGDDSDSFVAWLWDHLAESVDEYFSSNVEGTTIK
                SSLISSHEDKALVPLDSEYDKGRSDKSNGARRGRQWRSQPTNVSEIPPLLSSEVHKIH
                NYEKKDHKHRHNKRSPSPQAQSQRKRSRTDDSRNEQRASLRSALYETNKEAKPDVSRR
                LLQFAVRDALAISRPANSSTESSLKRLRSVVSTSTQNSSDPDPARKIRSVARVVNPMA
                TVMKAVAEAAEDAKKPKSGRSVFDRISHSTGSETLDQNMVLGEVSPKDEESRNFSQGQ
                EAVQLQYTQSLANNGVYFENMTTNDTGLNPDFSSDRGRLGSSVNVSHPSTYLGNRINN
                PNSLQHRLVDDSKRVKGTNYQNRLPEVATKHKTASFSGNIDTGKTVKLEEQRKIPDVG
                LQRYMDEGRLVSSEATTQLSTQKILGDTIGNGNVKPAASVKEESTTNKSVPGTLSTTR
                PLEDASSRTIFVANVHFGATKDSLSRHFNKFGEVLKAFIVTDPATGQPSGSAYIEFTR
                KEAAENALSLDGTSFMSRILKIVKGSNGQNQEAASSMSWSRGGRFTRAPSYFRGGAVR
                GRSVVRGGGRSMQWKRDSADTGNNNNVAPNNARSLTYVRAESKSDGVAND"
     gene       complement(14211. .15087)
                /gene="T2E22.6"
     mRNA       complement(<14211. .>15087)
                /gene="T2E22.6"
     repeat_region complement(14374. .14444)
                /rpt_family="AT_rich"
     CDS        complement(14605. .15087)
                /gene="T2E22.6"
                /codon_start=1
                /product="unknown protein; 15087-14605"
                /protein_id="AAG51008.1"
                /db_xref="GI:12321951"
                /translation="MAQRTEKEETEFKVLETLTTTTTTLCTNNCGVTANPATNNMCQK
                CFNASLVSAAAGVVESGSILKRSARSVNLRSSPAKVVIRPREIDAVKKRDQQIVNRCS
                GCRKKVGLTGFRCRCGELFCSEHRYSDRHDCSYDYKTAGREAIARENPVVKAAKMVKV
                "
     repeat_region complement(16197. .16239)
                /rpt_family="AT_rich"
     repeat_region 16366. .16406
                /rpt_family="(A)n"
     gene       16828. .18284
                /gene="T2E22.7"
                /note="contains Pfam profile: PF00481 Protein phosphatase
                2C"
     mRNA       join(<16828. .17125,17215. .17582,17733. .17969,
                18057. .>18284)
                /gene="T2E22.7"

Alignment Scores:
Pred. No.:          2.08e+04        Length:         105768
Score:              42.00           Matches:        6
Percent Similarity: 100.00%         Conservative:   0
Best Local Similarity: 100.00%      Mismatches:     0
Query Match:        100.00%         Indels:         0
DB:                 15              Gaps:           0

US-09-858-332G-28 (1-6) x AC069474 (1-105768)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      67663 CACAACCACAACCACAAC 67646

RESULT 582
AC011450/c
LOCUS       AC011450               106168 bp    DNA     linear   PRI 29-APR-2000
DEFINITION  Homo sapiens chromosome 19 clone CTC-30107, complete sequence.
ACCESSION   AC011450
VERSION     AC011450.4  GI:7670120
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 106168)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 106168)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 106168)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-2000) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Apr 29, 2000 this sequence version replaced gi:7021594.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.8% of Sequence;
            Estimated Total Number of Errors is 0.3.
            STS Content:
            SHGC-34723 G28380
            SHGC-58287 G38466.
FEATURES             Location/Qualifiers
     source          1. .106168
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="CTC-30107"
ORIGIN

Alignment Scores:
Pred. No.:          2.09e+04        Length:         106168
Score:              42.00           Matches:        6
Percent Similarity: 100.00%         Conservative:   0
Best Local Similarity: 100.00%      Mismatches:     0
Query Match:        100.00%         Indels:         0
DB:                 8               Gaps:           0

US-09-858-332G-28 (1-6) x AC011450 (1-106168)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      35605 CATAATCATAATCATAAT 35588

RESULT 583
AP007725
```

```
LOCUS       AP007725               106769 bp    DNA     linear   HTG 28-DEC-2004
DEFINITION  Lotus corniculatus var. japonicus chromosome 1 clone LjT43D23, ***
            SEQUENCING IN PROGRESS ***, 38 unordered pieces.
ACCESSION   AP007725
VERSION     AP007725.1  GI:56806035
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1
  AUTHORS   Kaneko,T., Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Structural Analysis of a Lotus japonicus Genome. XI. Sequence
            Features and Mapping of Nine hundred twenty-one TAC Clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 106769)
  AUTHORS   Sato,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-2004) Shusei Sato, Kazusa DNA Research Institute,
            Department of Plant Gene Research; 2-6-7 Kazusa-kamatari, Kisarazu,
            Chiba, 292-0818, Japan (E-mail:ssato@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/, Tel:81-438-52-3935(ex.2337),
            Fax:81-438-52-3934)
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 38 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1334: contig of 1334 bp in length
            *     1335     1434: gap of unknown length
            *     1435     2573: contig of 1139 bp in length
            *     2574     2673: gap of unknown length
            *     2674     3659: contig of 986 bp in length
            *     3660     3759: gap of unknown length
            *     3760     4757: contig of 998 bp in length
            *     4758     4857: gap of unknown length
            *     4858     6240: contig of 1383 bp in length
            *     6241     6340: gap of unknown length
            *     6341     7458: contig of 1118 bp in length
            *     7459     7558: gap of unknown length
            *     7559     9087: contig of 1529 bp in length
            *     9088     9187: gap of unknown length
            *     9188    10244: contig of 1057 bp in length
            *    10245    10344: gap of unknown length
            *    10345    11794: contig of 1450 bp in length
            *    11795    11894: gap of unknown length
            *    11895    13484: contig of 1590 bp in length
            *    13485    13584: gap of unknown length
            *    13585    14768: contig of 1184 bp in length
            *    14769    14868: gap of unknown length
            *    14869    16197: contig of 1329 bp in length
            *    16198    16297: gap of unknown length
            *    16298    18079: contig of 1782 bp in length
            *    18080    18179: gap of unknown length
            *    18180    19938: contig of 1759 bp in length
            *    19939    19948: gap of 10 bp
            *    19949    20036: contig of 88 bp in length
            *    20037    20136: gap of unknown length
            *    20137    21383: contig of 1247 bp in length
            *    21384    21483: gap of unknown length
            *    21484    23963: contig of 2480 bp in length
            *    23964    24063: gap of unknown length
            *    24064    25340: contig of 1277 bp in length
            *    25341    25440: gap of unknown length
            *    25441    28317: contig of 2877 bp in length
            *    28318    28417: gap of unknown length
            *    28418    31455: contig of 3038 bp in length
            *    31456    31555: gap of unknown length
            *    31556    34715: contig of 3160 bp in length
            *    34716    34815: gap of unknown length
            *    34816    37474: contig of 2659 bp in length
            *    37475    37574: gap of unknown length
            *    37575    40549: contig of 2975 bp in length
            *    40550    40649: gap of unknown length
            *    40650    43583: contig of 2934 bp in length
            *    43584    43683: gap of unknown length
            *    43684    46276: contig of 2593 bp in length
            *    46277    46376: gap of unknown length
            *    46377    50238: contig of 3862 bp in length
            *    50239    50338: gap of unknown length
            *    50339    53674: contig of 3336 bp in length
            *    53675    53774: gap of unknown length
            *    53775    56735: contig of 2961 bp in length
            *    56736    56835: gap of unknown length
            *    56836    60366: contig of 3531 bp in length
            *    60367    60466: gap of unknown length
            *    60467    64375: contig of 3909 bp in length
            *    64376    64475: gap of unknown length
            *    64476    68939: contig of 4464 bp in length
            *    68940    69039: gap of unknown length
            *    69040    72645: contig of 3606 bp in length
            *    72646    72745: gap of unknown length
            *    72746    76776: contig of 4031 bp in length
            *    76777    76876: gap of unknown length
            *    76877    81360: contig of 4484 bp in length
            *    81361    81460: gap of unknown length
            *    81461    87098: contig of 5638 bp in length
            *    87099    87198: gap of unknown length
            *    87199    92910: contig of 5712 bp in length
            *    92911    93010: gap of unknown length
            *    93011   101453: contig of 8443 bp in length
            *   101454   101553: gap of unknown length
            *   101554   106769: contig of 5216 bp in length.
FEATURES             Location/Qualifiers
     source          1. .106769
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="genomic DNA"
                     /variety="japonicus"
                     /db_xref="taxon:34305"
                     /chromosome="1"
                     /clone="LjT43D23"
                     /clone_lib="LjT library"
                     /note="TAC clone:TM0988, synonym:Lotus japonicus"
     gap             1335. .1434
                     /estimated_length=unknown
     gap             2574. .2673
                     /estimated_length=unknown
     gap             3660. .3759
                     /estimated_length=unknown
     gap             4758. .4857
                     /estimated_length=unknown
     gap             6241. .6340
                     /estimated_length=unknown
     gap             7459. .7558
                     /estimated_length=unknown
     gap             9088. .9187
                     /estimated_length=unknown
     gap             10245. .10344
                     /estimated_length=unknown
     gap             11795. .11894
                     /estimated_length=unknown
     gap             13485. .13584
                     /estimated_length=unknown
     gap             14769. .14868
                     /estimated_length=unknown
     gap             16198. .16297
                     /estimated_length=unknown
     gap             18080. .18179
                     /estimated_length=unknown
     gap             19939. .19948
                     /estimated_length=10
```

```
     gap             20037. .20136
                     /estimated_length=unknown
     gap             21384. .21483
                     /estimated_length=unknown
     gap             23964. .24063
                     /estimated_length=unknown
     gap             25341. .25440
                     /estimated_length=unknown
     gap             28318. .28417
                     /estimated_length=unknown
     gap             31456. .31555
                     /estimated_length=unknown
     gap             34716. .34815
                     /estimated_length=unknown
     gap             37475. .37574
                     /estimated_length=unknown
     gap             40550. .40649
                     /estimated_length=unknown
     gap             43584. .43683
                     /estimated_length=unknown
     gap             46277. .46376
                     /estimated_length=unknown
     gap             50239. .50338
                     /estimated_length=unknown
     gap             53675. .53774
                     /estimated_length=unknown
     gap             56736. .56835
                     /estimated_length=unknown
     gap             60367. .60466
                     /estimated_length=unknown
     gap             64376. .64475
                     /estimated_length=unknown
     gap             68940. .69039
                     /estimated_length=unknown
     gap             72646. .72745
                     /estimated_length=unknown
     gap             76777. .76876
                     /estimated_length=unknown
     gap             81361. .81460
                     /estimated_length=unknown
     gap             87099. .87198
                     /estimated_length=unknown
     gap             92911. .93010
                     /estimated_length=unknown
     gap             101454. .101553
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              2.1e+04         Length:         106769
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AP007725 (1-106769)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      35800 CATAACCATAATCATAAC 35817

RESULT 584
AL935319/c
LOCUS       AL935319              106806 bp    DNA     linear   VRT 08-APR-2003
DEFINITION  Zebrafish DNA sequence from clone DKEY-56L2 in linkage group 18,
            complete sequence.
ACCESSION   AL935319
VERSION     AL935319.8  GI:29691818
KEYWORDS    HTG.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 106806)
  AUTHORS   Beasley,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 9, 2003 this sequence version replaced gi:28550591.
            ------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep Repeat names
            beginning 'Dr' were identified by the Recon repeat discovery system
            (Zhirong Bao and Sean Eddy, submitted), and those beginning 'drr'
            were identified by Rick Waterman (Stephen Johnson lab, WashU).  For
            further information see http://www/Projects/D_rerio/fishmask.shtml
            DKEY-56L2 is from a Zebrafish BAC library
            VECTOR: pIndigoBAC-5.
FEATURES             Location/Qualifiers
     source          1. .106806
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-56L2"
                     /clone_lib="DanioKey"
ORIGIN

Alignment Scores:
Pred. No.:              2.1e+04         Length:         106806
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     5               Gaps:           0

US-09-858-332G-28 (1-6) x AL935319 (1-106806)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      25173 CATAATCATAATCATAAT 25156

RESULT 585
AC024337/c
LOCUS       AC024337              106945 bp    DNA     linear   PRI 24-MAR-2002
DEFINITION  Homo sapiens chromosome 15, clone RP11-236L14, complete sequence.
ACCESSION   AC024337
VERSION     AC024337.8  GI:19698714
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 106945)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-236L14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 106945)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
            Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
            McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
            Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
            Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 106945)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 106945)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 24, 2002 this sequence version replaced gi:19683752.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5452
                Center clone name: 236_L_14
            --------------

            Only the last 106.9 kilobases of this clone are being submitted.
            The remainder overlaps accession number AC018574 [WICGR project
            L2544].
FEATURES             Location/Qualifiers
     source          1. .106945
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-236L14"
                     /clone_lib="RPCI-11 Human Male BAC"
     unsure          21. .54
                     /note="<30 qual SNGL region"
     unsure          98. .104
                     /note="<30 qual SNGL region"
     unsure          135. .139
                     /note="<30 qual SNGL region"
     unsure          175. .179
                     /note="<30 qual SNGL region"
     repeat_region   605. .717
                     /rpt_family="L2"
     repeat_region   complement(3603. .4466)
                     /rpt_family="L2"
     repeat_region   4467. .4599
                     /rpt_family="FLAM_A"
     repeat_region   complement(4600. .5412)
                     /rpt_family="L2"
     repeat_region   complement(5413. .5772)
                     /rpt_family="THE1B"
     repeat_region   complement(5773. .5909)
                     /rpt_family="L2"
     repeat_region   7701. .7774
                     /rpt_family="(TA)n"
     repeat_region   7787. .7841
                     /rpt_family="AT_rich"
     unsure          8831. .8835
                     /note="<30 qual SNGL region"
     repeat_region   9968. .10135
                     /rpt_family="AluSg/x"
     unsure          10033. .10038
                     /note="<30 qual SNGL region"
     unsure          10089. .10094
                     /note="<30 qual SNGL region"
     unsure          10115. .10128
                     /note="<30 qual SNGL region"
     repeat_region   complement(10385. .10500)
                     /rpt_family="L2"
     repeat_region   complement(10962. .11170)
```

```
                     /rpt_family="MIR"
     repeat_region   complement(12488. .12551)
                     /rpt_family="L2"
     repeat_region   13096. .13253
                     /rpt_family="MIR"
     repeat_region   13272. .13294
                     /rpt_family="AT_rich"
     repeat_region   14757. .14784
                     /rpt_family="AT_rich"
     repeat_region   17770. .17816
                     /rpt_family="(CA)n"
     repeat_region   18116. .18161
                     /rpt_family="CT-rich"
     repeat_region   18193. .18409
                     /rpt_family="MER102b"
     repeat_region   complement(19342. .19634)
                     /rpt_family="AluJb"
     repeat_region   20332. .20368
                     /rpt_family="(CA)n"
     repeat_region   20619. .20778
                     /rpt_family="MIR"
     repeat_region   complement(22035. .22339)
                     /rpt_family="AluY"
     repeat_region   23152. .23191
                     /rpt_family="CT-rich"
     repeat_region   complement(23415. .24545)
                     /rpt_family="L1MC1"
     repeat_region   24949. .25259
                     /rpt_family="AluJo"
     repeat_region   25466. .25494
                     /rpt_family="AT_rich"
     repeat_region   complement(25771. .26264)
                     /rpt_family="MLT1J"
     repeat_region   complement(27708. .27754)
                     /rpt_family="MER69"
     repeat_region   27755. .28120
                     /rpt_family="THE1B"
     repeat_region   complement(28121. .28184)
                     /rpt_family="MER69"
     repeat_region   28614. .28907
                     /rpt_family="AluSx"
     repeat_region   29092. .29253
                     /rpt_family="MER5A"
     repeat_region   29373. .29646
                     /rpt_family="AluSq"
     repeat_region   complement(30322. .30456)

Alignment Scores:
Pred. No.:              2.1e+04         Length:         106945
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     8               Gaps:           0

US-09-858-332G-28 (1-6) x AC024337 (1-106945)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db       2489 CACAACCACAACCACAAC 2472

RESULT 586
AC006551/c
LOCUS       AC006551               107200 bp    DNA     linear   PLN 11-JUN-2001
DEFINITION  Arabidopsis thaliana chromosome I BAC F12K8 genomic sequence,
            complete sequence.
ACCESSION   AC006551
VERSION     AC006551.6  GI:6056182
KEYWORDS    HTG.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 107200)
  AUTHORS   Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
            Altafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
            Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,
            Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vysotskaia,V.S.,
            Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 107200)
  AUTHORS   Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
            Altafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
            Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,
            Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V.,
            Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-FEB-1999) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
REFERENCE   3  (bases 1 to 107200)
  AUTHORS   Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
            Altafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
            Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,
            Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V.,
            Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-OCT-1999) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
REFERENCE   4  (bases 1 to 107200)
  AUTHORS   Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
            Altafi,H., Araujo,R., Huizar,L., Rowley,D., Brooks,S., Buehler,E.,
            Chao,Q., Dunn,P., Gonzalez,A., Khan,S., Kremenetskaia,I., Kim,C.,
            Lenz,C., Li,J., Liu,S., Luros,S., Schwartz,J., Shinn,P.,
            Toriumi,M., Vyotskaia,V., Yu,G., Ecker,J., Theologis,A. and
            Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-DEC-1999) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
COMMENT     On Oct 16, 1999 this sequence version replaced gi:5881526.
            Bases 1-12905 of clone F12K8 overlap with bases 70259-83163 of TAMU
            BAC clone T22J18, gb|AC003979.
FEATURES             Location/Qualifiers
     source          1. .107200
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /chromosome="I"
                     /clone="F12K8"
     gene            complement(816. .1713)
                     /gene="F12K8.1"
     CDS             complement(join(816. .1326,1407. .1536,1614. .1713))
                     /gene="F12K8.1"
                     /note="Location of ests 250F14T7 (gb|AA598046), 250E15T7
                     (gb|AA598042), 110A14T7 (gb|T42000 5263), and 1B4T7P
                     (T04118). 58% identical to DNA-binding protein
                     [Arabidopsis thaliana] (gb|U26936)."
                     /codon_start=1
                     /product="Putative DNA binding protein"
                     /protein_id="AAF18515.1"
                     /db_xref="GI:6587826"
                     /translation="MNKGAWTKEEDQLLVDYIRKHGEGCWRSLPRAAGLQRCGKSCRL
                     RWMNYLRPDLKRGNFTEEEDELIIKLHSLLGNKWSLIAGRLPGRTDNEIKNYWNTHIK
                     RKLLSRGIDPNSHRLINESVVSPSSLQNDVVETIHLDFSGPVKPEPVREEIGMVNNCE
                     SSGTTSEKDYGNEEDWVLNLELSVGPSYRYESTRKVSVVDSAESTRRWGSELFGAHES
                     DAVCLCCRIGLFRNESCRNCRVSDVRTH"
     gene            complement(3868. .4548)
                     /gene="F12K8.2"
     CDS             complement(join(3868. .3955,4062. .4108,4259. .4318,
                     4411. .4548))
                     /gene="F12K8.2"
                     /note="Unknown protein; Location of ests 89E10XP 3'
                     (gb|AI100452), 207C19T7 (gb|N37611), and 89E10T7
```

```
                (gb|T20896)"
                /codon_start=1
                /protein_id="AAF18516.1"
                /db_xref="GI:6587827"
                /translation="MLKFPLCCSELSLEPKILAFESQLVSGRRDTIKVSAGKIGNFSL
                GSIFKSCETCGAKGAIECPGCKGTGKNKKNGNMFERWKCFDCQGFGMKSCPKCGKGGL
                TPEQRGER"
gene            5248. .10146
                /gene="F12K8.3"
CDS             join(5248. .5379,5566. .5770,5860. .6025,6152. .6374,
                6452. .6528,6752. .6939,7064. .7161,7227. .7384,7457. .7591,
                7713. .8112,8641. .8726,8916. .9128,9317. .9502,9690. .9801,
                9895. .10146)
                /gene="F12K8.3"
                /note="Hypothetical protein"
                /codon_start=1
                /protein_id="AAF18517.1"
                /db_xref="GI:6587828"
                /translation="MAKSENSTTSTFSSFANKIQPSNDAESDPDSYALEKFKLYETRA
                RFYLVGSDRNKRFFRVLKIDRSEPSELNISEDPVVYSPQEIKSLLQRIAEGNRATGGL
                AFVAKVYGIAGCAKFMESYYLVLVTKRRQIGCICGHAIYAIDESQMISVPHATIQSDV
                ANSKTELRYKKLLSSVDLTKDFFYSYTYPIMQSLQKNVLSSGEEGMPYDNIFVWNSYL
                TQPIRSRCNNTIWTLALVHGHFKQIRLSIYGRDFSVTLVSRRSRHFAGTRYLKRGVND
                RGRVANDVETEQLVLDDEAGSCKGKMSSVVQMRGSIPLFWSQEASRFSPKPDIFLQRY
                DPTYESTKMHFEDLVNRYGNPIIVLNLIKTNIHGNIQTVEKRPREMVLRREFANAVGY
                LNSIFREENHLKFIHWDFHKFAKSKSANVLAVLGAVASEALDLTGLYFSGKPKIVKKK
                ASQLSHANTARDKYCREPSLRDLRAYSAELSRGESANDILSALANREKEMKLTQQKKD
                EGTNSSAPRYQSGVLRTNCIDCLDRTNVAQYAYGLAALGRQLHAMGLSDTPKIDPDSS
                IAAALMDMYQSMGDALAQQYGGSAAHNTEGKPALWELDSDYYLHVSGIGDDIFPDIGV
                QSIAKPMSGIGVNLAPVPAFRDDFSRKKLTSFDKLIEQTCSSIKNVRLCSETDQRPGG
                NTGSTGVAPDAAEIQLKSPNWLFGSRKPEESSSATKSGADDSEKGVTSTERVNDFCNL
                DWLSKSDRHQGDIFQRYLSITSTNEANGWYGGTLLGDQDENSEIYRHYAQFCQCPAME
                PFENDHEFEQNFAEVLRMNTIDVMDIEEEETEMESDFNEYTQIGSDLGIIPMQCKHFA
                SDPCWLARWLVGDDKVPKVI"
gene            complement(10468. .13557)
                /gene="F12K8.4"
CDS             complement(10468. .13557)
                /gene="F12K8.4"
                /note="Last 369 amino acids are 61% identical to
                phosphoribosylanthranilate transferase [Pisum sativum]
                (gb|BAA13032)."
                /codon_start=1
                /product="Highly similar to phosphoribosylanthranilate
                transferase"
                /protein_id="AAF18518.1"
                /db_xref="GI:6587829"
                /translation="MNKLVVEIVDASDLMPKDGQGSASPFVEVEFDEQRQRTQTRFKD
                LNPQWNEKLVFNVGDLKRLNNKTVDVTVYDDRRDNQPGKFLGRVKIAGAVVPLSESES
                GVQRYPLDKRGLFSNIKGDIALRIYAAPIDGGDFVSPPPDFAEKVMKEDKRFESQEFQ
                FQNQNQNQNHYEQFEDEINNMETLKPTKKKEKESRTFHSIGAHAGGGGGAPPMSQAKQ
                AYPPPPNQPEFRSDFMRAPGPPTGAVMQMQPPRQQNPEFQLIETSPPLAARMRQSYYY
                RSSGDKTSSTYDLVEQMHYLYVSVVKARDLPVMDVSGSLDPYVEVKLGNYKGLTKHLE
                KNSNPIWKQIFAFSKERLQSNLLEVTVKDKDLLTKDDFVGRVHIDLTEVPLRVPPDSP
                LAPQWYRLEDKKGMKTNRGEIMLAVWMGTQADESFPDAWHSDAHRVSHSNLSNTRSKV
                YFSPKLYYLRIHVMEAQDLVPSDKGRVPDAIVKIQAGNQMRATRTPQMRTMNPQWHEE
                LMFVVSEPFEDMVIVSVDDRIGPGKDEILGRVFIPVRDVPVRQEVGKMPDPRWFNLQR
                HSMSMEEENEKRKEKFSSKILLRVCIEAGYHVLDESTHFSSDLQPSSKHLRKPSIGIL
                ELGILSARNLMPMKGKDGRMTDPYCVAKYGNKWVRTRTLLDALAPKWNEQYTWEVHDP
                CTVITIGVFDNSHVNDGGDFKDQRIGKVRVRLSTLETDRVYTHFYPLLVLTPGGLKKN
                GELQLALRYTCTGFVNMMAQYGRPLLPKMHYIQPIPVRHIDLLRHQAMQIVATRLSRS
                EPPLRRREVVEYMLDVDYHMFSLRRSKANFSRIMSLLSSVTLVCKWFNDICTWRNPITT
                CLVHVLFLILVCYPELILPTVFLYLFVIGMWNYRYRPRHPPHMDARVSQADNAHPDEL
                DEEFDTFPTSRPADIVRMRYDRLRSVGGRVQTVVGDLATQGERIQALLSWRDPRATAL
                FIVFALIWAVFIYVTPFQVIAIIIGLFMLRHPRFRSRMPSVPANFFKRLPAKSDMLL"
gene            18848. .20727
                /gene="F12K8.5"
CDS             join(18848. .19732,20455. .20727)
                /gene="F12K8.5"
                /note="36% identical to seed maturation protein PM27
                [Glycine max] (gb|AAD30426)."
                /codon_start=1
                /product="Similar to seed maturation protein PM27"
                /protein_id="AAF18519.1"
                /db_xref="GI:6587830"
                /translation="MTKLLALSLVLSALFTVGHCTGSEDDIIVRDGHRVVVVEYDRDG
                ETNTRVLISPPGKEVDQKSRDEGEVFGNEKRETASSLPEEEEREHHATPGELICDAIG
                KCKHKLGTVLGRVKDRTASDLSDETPEMTVAREALEVEEKVSWKAREARGKVNERATK
                KAHRVQKVLEKVQIAVRGIGTVVATALGLTKIGSVVGIVGIAAAYGMCVWVTFVSGYV
                LASVLGEQQFGVVQSKMYPVYFKAVSVGILVGLLGHVIGRRRKVFTDAVDMWQSVNLL
                SSILMVEANASFVYTRATKAMFELIKAEKEDGRGFDTSDQSQSSESAGRTRGKKKVTE
                KTDEDVVKQRLTKLSERMRKLNAYSSRLNLLTLMSLTWHFVYLGYRLSLTC"
gene            21976. .22641
                /gene="F12K8.6"
CDS             join(21976. .22094,22206. .22641)
                /gene="F12K8.6"
                /note="Hypothetical protein"
                /codon_start=1
                /protein_id="AAF18539.1"
                /db_xref="GI:6587850"
                /translation="MMGLRFWLENFGFKLIGGKISGLRFWRKITDLRFWRENLGFTFL
                AEKSRVYVFDGKIMGLRFGRKNKGFTFLAEKSRVYVSGGKITGLRFWRENHGVTFLAE
                KSRVYVFGGKITGLRFWRENYGFTFLAGKLWIYVFGGKITGLRFWRENHGFMFFGGKI
                TSLFFLNCIACIFKKFGKILILLN"
gene            complement(24033. .26186)
                /gene="F12K8.7"
CDS             complement(join(24033. .24315,25922. .25990,26161. .26186))
                /gene="F12K8.7"
                /note="Unknown protein; Location of ests F13558
                (gb|F13558) and Z37169 (gb|Z37169)"
                /codon_start=1
                /protein_id="AAF18520.1"
                /db_xref="GI:6587831"
                /translation="MTMAPKRCFKTNESTKSVRVACRSLRSPSVRRLSELPVEKQTKY
                MMDQKDLMNKMIQDAEKKLEKEKMHTRAMKLGLMAGSNDLITDTDCSEELARAADVVD
                KKLKAIRERIKAVEAGAPIIKRD"
gene            29462. .31415
                /gene="F12K8.8"
CDS             join(29462. .29585,29667. .29884,29981. .30540,30620. .31415)
                /gene="F12K8.8"
                /note="41% identical to LeOPT1 [Lycopersicon esculentum]
                (gb|AF016713)."
                /codon_start=1
                /product="Similar to LeOPT1 [Lycopersicon esculentum]"
                /protein_id="AAF18521.1"
                /db_xref="GI:6587832"
                /translation="MKIPEEEVALLEDYVSDSVDHRGFPAGKSSTGGWRSAWFIIGVE
                VAERFAYFGIACNLITYLTGPLGQSTAKAAVNVNTWSGTASILPILGAFVADAYLGRY
```

```
Alignment Scores:
Pred. No.:              2.11e+04        Length:         107200
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x AC006551 (1-107200)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      62971 CATAATCATAATCATAAT 62954

RESULT 587
AC003070/c
LOCUS       AC003070              107217 bp    DNA     linear   PRI 31-DEC-2002
DEFINITION  Homo sapiens chromosome 17, clone CTB-39G8, complete sequence.
ACCESSION   AC003070
VERSION     AC003070.2  GI:27436829
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 107217)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
```

```
  TITLE     Homo sapiens chromosome 17, clone CTB-39G8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 107217)
  AUTHORS   Birren,B., Fasman,K., McKernan,K., Munro,C., Nusbaum,C.,
            Richardson,P., Lander,E., Baldwin,J., Barna,N., Cantu,C., Chang,A.,
            Cooke,P., Daly,M.J., Devon,K., Dewar,K., DuRette,B., Forrest,C.,
            Gage,D., Gensheimer,S., Geraigery,K., Gilmartin,T., Hagos,B.,
            Halphen,I., Harris,K., Howland,J.C., Huang,J., Hui,L., Jacotot,L.,
            Kirby,A., Lane,M., MacKenzie,J., Marquis,N., McDermott,J.,
            Molla,M., Morrow,J., Nachman,A., Naylor,J., O'Connor,T., Olotu,A.,
            Peterson,K., Roberts,D., Rollins,G., Sarnaik,A., Shiu,P., Shyam,R.,
            Stilwell,J., Stone,C., Strickland,C., Sydney,K., Tang,L.,
            Zemtseva,I. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1997) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 107217)
  AUTHORS   Birren,B., Fasman,K., McKernan,K., Nusbaum,C., Richardson,P.,
            Lander,E., Allen,N., Baker,J., Baldwin,J., Barna,N., Beckerly,R.,
            Boutwell,C., Byrne,S., Cantu,C., Castle,A., Cooke,P., Daly,M.J.,
            Depayre,E., Devon,K., Dewar,K., DuRette,B., Etemadi,S.,
            Ferreira,P., Forrest,C., Gage,D., Gardyna,S., Gensheimer,S.,
            Geraigery,K., Gilmartin,T., Gray,D., Hagos,B., Harris,K.,
            Horton,L., Howland,J.C., Hui,L., Jacotot,L., Linton,L.,
            MacKenzie,J., Marquis,N., McEwan,P., McGurk,A., Meldrim,J.,
            Molla,M., Morris,W., Morrow,J., Nachman,A., Naylor,J., O'Connor,T.,
            Pavlin,B., Peterson,K., Ranganath,S., Riley,R., Roberts,D.,
            Rollins,G., Rossello,R., Roy,A., Shyam,R., Soohoo,S., Stilwell,J.,
            Stone,C., Strickland,C., Sydney,K., Tang,L., Vassiliev,H., Vo,A.,
            Wagner,A., Wheeler,J., Wu,Y., Ye,W.J., Zemtseva,I., Zhao,J. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-FEB-1998) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 107217)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-DEC-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Dec 31, 2002 this sequence version replaced gi:2865212.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L266
                Center clone name: 39_G_8
            --------------

            Only the last 107.2 kilobases of this clone are being submitted.
            The remainder overlaps accession number AC091132 [WICGR project
            L11996].
```

```
FEATURES             Location/Qualifiers
     source          1. .107217
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="CTB-39G8"
                     /clone_lib="CITB Human BAC"
     repeat_region   1. .273
                     /rpt_family="AluJo"
     repeat_region   complement(594. .892)
                     /rpt_family="AluSx"
     repeat_region   complement(1170. .1338)
                     /rpt_family="MIR"
     repeat_region   1452. .1748
                     /rpt_family="AluSx"
     repeat_region   1768. .1898
                     /rpt_family="MER5B"
     repeat_region   2889. .2920
                     /rpt_family="(TG)n"
     repeat_region   complement(5070. .5242)
                     /rpt_family="L3"
     repeat_region   complement(5722. .6117)
                     /rpt_family="L2"
     repeat_region   6271. .6506
                     /rpt_family="L2"
     repeat_region   6903. .7213
                     /rpt_family="AluSx"
     repeat_region   complement(7342. .7558)
                     /rpt_family="L2"
     repeat_region   7960. .8055
                     /rpt_family="AluSg/x"
     repeat_region   8176. .8204
                     /rpt_family="AT_rich"
     repeat_region   complement(8227. .8373)
                     /rpt_family="MIR"
     repeat_region   complement(8504. .8597)
                     /rpt_family="L1ME4A"
     repeat_region   complement(10738. .10924)
                     /rpt_family="MIR"
     repeat_region   complement(11137. .11438)
                     /rpt_family="AluSx"
     repeat_region   12061. .12182
                     /rpt_family="L3"
     repeat_region   13277. .13347
                     /rpt_family="MER117"
     repeat_region   complement(13348. .13656)
                     /rpt_family="AluJo"
     repeat_region   13657. .13741
                     /rpt_family="MER117"
     repeat_region   13761. .13804
                     /rpt_family="(TTA)n"
     repeat_region   complement(13806. .14083)
                     /rpt_family="AluSx"
     repeat_region   complement(14531. .14831)
                     /rpt_family="AluJo"
     repeat_region   complement(15092. .15405)
                     /rpt_family="AluY"
     repeat_region   complement(15960. .16262)
                     /rpt_family="AluSg"
     repeat_region   16394. .16495
                     /rpt_family="MIR"
     repeat_region   17043. .17379
                     /rpt_family="L3"
     repeat_region   complement(19972. .20042)
                     /rpt_family="MIR"
     repeat_region   complement(20497. .20573)
                     /rpt_family="MIR"
     repeat_region   complement(20681. .21202)
                     /rpt_family="MER115"
     repeat_region   complement(21398. .21651)
                     /rpt_family="L2"
```

```
     repeat_region   complement(21783. .21917)
                     /rpt_family="MIR3"
     repeat_region   21968. .22045
                     /rpt_family="L2"
     repeat_region   complement(22390. .22688)
                     /rpt_family="AluJb"
     repeat_region   23207. .23427
                     /rpt_family="L2"
     repeat_region   complement(23434. .23482)
                     /rpt_family="L1MB8"
     repeat_region   complement(23483. .23795)
                     /rpt_family="AluJb"
     repeat_region   complement(23796. .23892)
                     /rpt_family="L1MB8"
     repeat_region   complement(23893. .24204)
                     /rpt_family="AluSg1"
     repeat_region   complement(24205. .24282)
                     /rpt_family="L1MB8"
     repeat_region   complement(24283. .24635)
                     /rpt_family="MLT1A2"
     repeat_region   complement(24636. .24810)
                     /rpt_family="L1MB8"
     repeat_region   complement(25095. .25453)
                     /rpt_family="LTR16A"
     repeat_region   25474. .26467
                     /rpt_family="L2"
     repeat_region   26482. .26709
                     /rpt_family="MIR"
     repeat_region   26662. .27115
                     /rpt_family="L2"
     repeat_region   27118. .27412
                     /rpt_family="AluY"
     repeat_region   27436. .27621
                     /rpt_family="MIR"
     repeat_region   complement(28227. .28771)
                     /rpt_family="L1MD2"
     repeat_region   30191. .30430
                     /rpt_family="L2"
     repeat_region   30430. .30544
                     /rpt_family="L2"
     repeat_region   30545. .30586
                     /rpt_family="(TGAA)n"

Alignment Scores:
Pred. No.:              2.11e+04        Length:         107217
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     8               Gaps:           0

US-09-858-332G-28 (1-6) x AC003070 (1-107217)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      13796 CATAATCATAATCATAAT 13779

RESULT 588
AP005290
LOCUS       AP005290              107284 bp    DNA     linear   PLN 13-JUL-2004
DEFINITION  Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 2,
            BAC clone:OJ1282_E10.
ACCESSION   AP005290
VERSION     AP005290.2  GI:38198228
KEYWORDS    .
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Sasaki,T., Matsumoto,T. and Katayose,Y.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 2, BAC
            clone:OJ1282_E10
  JOURNAL   Published Only in Database (2002)
REFERENCE   2  (bases 1 to 107284)
  AUTHORS   Sasaki,T., Matsumoto,T. and Katayose,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAY-2002) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     On Nov 6, 2003 this sequence version replaced gi:21280340.
            Genes were predicted from the integrated results of the following:
            GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH
            (http://www.softberry.com/), GeneMark.hmm
            (http://opal.biology.gatech.edu/GeneMark/), GlimmerM
            (http://www.tigr.org/tdb/glimmerm/glmr_form.html), RiceHMM
            (http://rgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor
            (http://bioinformatics.iastate.edu/cgi-bin/sp.cgi), sim4
            (http://globin.cse.psu.edu/html/docs/sim4.html), gap2
            (http://www.tigr.org/software/glimmerm/), BLASTN and BLASTX. The
            genomic sequence was searched against NCBI NonRedundant Protein
            database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA
            sequence database at RGP or DDBJ. Protein homologies of the coding
            regions were searched against NCBI NonRedundant Protein database
            with BLASTP. ESTs represent the identified cDNA sequences using
            BLASTN with the corresponding DDBJ accession no. and RGP clone ID.
            Full-length cDNAs represent the identified cDNA sequences using
            BLASTN with the corresponding DDBJ accession no.
            A gene with identity or significant homology to a protein is
            classified based on the protein name to indicate the homology level
            such as same name, 'putative-' and '-like protein'.  A gene without
            significant homology to any protein but with full-length cDNA or
            EST homology (covering almost the entire length of partial
            sequence) is classified as an 'unknown' protein. A gene predicted
            by two or more gene prediction programs is classified as a
            'hypothetical' protein according to IRGSP standard. A gene
            predicted by a single gene prediction program is also classified as
            a probable 'hypothetical' protein and is included as a
            miscellaneous feature of the sequence.
            The orientation of the sequence is from M13rev to -21M13 of the BAC
            clone. This sequence of OJ1282_E10 clone has an overlap with
            OJ1149_C12 (DDBJ: AP004082) clone at 5' end and with P0264G11
            (DDBJ: AP006187) at 3' end. The sequence was generated by combining
            Monsanto and RGP-Japan sequencing data. Detailed information on
            overlap and assembly quality together with annotation of this entry
            is available at
            http://rgp.dna.affrc.go.jp/GenomeSeq.html.
FEATURES            Location/Qualifiers
     source         1. .107284
                    /organism="Oryza sativa (japonica cultivar-group)"
                    /mol_type="genomic DNA"
                    /cultivar="Nipponbare"
                    /db_xref="taxon:39947"
                    /chromosome="2"
                    /clone="OJ1282_E10"
     gene           13987. .15579
                    /gene="OJ1282_E10.1"
     mRNA           13987. .15355
                    /gene="OJ1282_E10.1"
                    /note="supported by full-length cDNA(s): AK072277"
     mRNA           13987. .15355
                    /gene="OJ1282_E10.1"
                    /note="supported by full-length cDNA(s): AK100140"
     mRNA           14055. .15579
                    /gene="OJ1282_E10.1"
                    /note="supported by full-length cDNA(s): AK058645"
     CDS            14134. .15189
                    /gene="OJ1282_E10.1"
                    /note="contains EST(s): C26049(C11540)
                    contains full-length cDNA(s): AK100140,AK072277,AK058645"
                    /codon_start=1
                    /product="PRLI-interacting factor A-like"
                    /protein_id="BAD28447.1"
```

```
                     /db_xref="GI:50251409"
                     /translation="MDPNQMMGRSFPMWAAADDPSAQPPPFLPPPNRGWKRKNPSTAA
                     GGYQPPALGDLQVQNRAKARRWFKNQSGGGGGGGGPRKHFFPRPKAAAPRNTTSFIIR
                     AKRAGGIASLVSPSPVTPAVLPTPRLSPSREGLSDMAQAQWGVDGYGSMKGLIRLRTS
                     PHPANPSDDDDDANSSGSDVEEHVEVERRLDHDLSRFEMVYPARGDGVFFEEEDDYEY
                     DQEAHVARLEEENLTLKERLFLMEQELGDMRRRLEALESRFAGTDGMPIENSDAVAVD
                     HDKEMAHPILVQRDDLRTEIAADAAGSDKASAQDNAEEVADAAVSEKIVQDNADVVVA
                     GSERTGEEIGVDASEMD"
gene                 join(16106. .16380,16460. .16683,19741. .19821,19921. .20009,
                     20129. .20198,20305. .20343,20444. .20594,20692. .20781,
                     21095. .21399)
                     /gene="OJ1282_E10.2-1"
mRNA                 join(16106. .16380,16460. .16683,19741. .19821,19921. .20009,
                     20129. .20198,20305. .20343,20444. .20594,20692. .20781,
                     21095. .21399)
                     /gene="OJ1282_E10.2-1"
                     /note="supported by full-length cDNA(s): AK099439"
gene                 join(16168. .16380,16460. .16683,19745. .19821,19921. .20009,
                     20129. .20343,20444. .20594,20692. .20781,21095. .21399)
                     /gene="OJ1282_E10.2-2"
mRNA                 join(16168. .16380,16460. .16683,19745. .19821,19921. .20009,
                     20129. .20343,20444. .20594,20692. .20781,21095. .21399)
                     /gene="OJ1282_E10.2-2"
                     /note="supported by full-length cDNA(s): AK099942"
CDS                  join(16472. .16683,19741. .19821,19921. .20009,20129. .20198,
                     20305. .20343,20444. .20594,20692. .20781,21095. .21193)
                     /gene="OJ1282_E10.2-2"
                     /note="contains EST(s): D46902(S11855),AU096009(S11855)
                     contains full-length cDNA(s): AK099439
                     metal tolerance protein-like"
                     /codon_start=1
                     /product="cation efflux family protein-like"
                     /protein_id="BAD28448.1"
                     /db_xref="GI:50251410"
                     /translation="MLSISPLVVESSLSPCLPWLPPEPNLITSTPTGASLPQFTPLFF
                     NLPLPLIHCCLYSAIKDWRSLLPLQMPQLFLLFLSFSLAVEALHAFMQDESEHKHYLI
                     VSAVTNLLVNLLGVWFFRSYARVNIVYRKAEDMNYHSVCLHVLADSVRSAGLILASWF
                     LSLGVENAEVLCLGIVSVAVFMLVLPLFKATGNILLQIAPGNVPPSALTKCFRQITAC
                     EDVSEVCQGRFWELVPGHAVGSLDIRVKNGGDCQSVLDYVHGLYQDLGIQDLTIQTDE
                     "
CDS                  join(19799. .19821,19921. .20009,20129. .20283)
                     /gene="OJ1282_E10.2-2"
                     /note="contains full-length cDNA(s): AK099942
                     metal tolerance protein-like"
                     /codon_start=1
                     /product="cation efflux family protein-like"
                     /protein_id="BAD28449.1"
                     /db_xref="GI:50251411"
                     /translation="MQDESEHKHYLIVSAVTNLLVNLLGVWFFRSYARVNIVYRKAED
                     MNYHSVCLHVLADSVRRYEILSCHCASSLIIFALSVVTKLVIFS"
gene                 complement(21642. .27218)
                     /gene="OJ1282_E10.3"
mRNA                 complement(join(21642. .21917,22155. .22325,22396. .22434,
                     22533. .22595,22734. .22827,22915. .23005,23091. .23152,
                     23255. .23324,23412. .23480,23567. .23633,24567. .24629,
                     24712. .24757,25088. .25164,25236. .25270,26026. .26088,
                     26178. .26265,26573. .26623,26747. .27218))
                     /gene="OJ1282_E10.3"
                     /note="supported by full-length cDNA(s): AK101526"
CDS                  complement(join(21807. .21917,22155. .22325,22396. .22434,
                     22533. .22595,22734. .22827,22915. .23005,23091. .23152,
                     23255. .23324,23412. .23480,23567. .23633,24567. .24629,
                     24712. .24757,25088. .25164,25236. .25270,26026. .26088,
                     26178. .26265,26573. .26623,26747. .26815))
                     /gene="OJ1282_E10.3"
                     /note="contains EST(s): C96708(C10601),D22247(C10601)
                     contains full-length cDNA(s): AK101526,AK058826"
                     /codon_start=1
                     /product="N-acetylglucosaminyltransferase I-like protein"
                     /protein_id="BAD28450.1"
                     /db_xref="GI:50251412"
                     /translation="MARSPCDLRILLVAAAAAFIYIQVRLFSTQSHYADRLAQAEKSE
                     NQCTSQLRSLIDQVSSQQEKIVALEEMKIRQDEERVHLKILIQDLEKRSVQTLVNNNV
                     APVAAVVVMACNRPDYLQRTVESILKYQTSVASKFPLFISQDGINGEVKKKALSYNEI
                     TYMQHLDLEPVRTERPGELIAYYKIAKHYKWALDELFIKHNFARVIILEDDMEIAPDF
                     FDYFEAAAKLLDNDKTIMAVSSWNDNGQKQFVYDPKALYRSDFFPGLGWMLTKPTWIE
                     LSPKWPKAYWDDWVRLKEVHRDRQFIRPEVCRTYNFGEHGSSMGQFFRQYLEPIKLND
                     AHIKWNSEDLSYLKEDKFLIQFGKDVASATPLHGSDAALKAHNMDADVRIQYNDQEDF
                     ERIARQFGIFEEWKDGIPRTAYKGVVVFRYKSSRRRIYLVGPDSLSQLRV"
gene                 complement(27358. .27570)
                     /gene="OJ1282_E10.4"
misc_feature         complement(27358. .27570)
                     /gene="OJ1282_E10.4"
                     /note="hypothetical ORF
                     predicted by GlimmerM
                     this category is not included in IRGSP standard"
gene                 28407. .29252
                     /gene="OJ1282_E10.5"
misc_feature         28407. .29252
                     /gene="OJ1282_E10.5"
                     /note="probably inactive due to including stop codon(s) in
                     CDS
                     pseudogene, heterogeneous nuclear ribonucleoprotein
                     (hnRNP)"
                     /pseudo
gene                 30725. .32010
                     /gene="OJ1282_E10.6"
mRNA                 30725. .32010
                     /gene="OJ1282_E10.6"
                     /note="supported by full-length cDNA(s): AK107147"
CDS                  30731. .31837
                     /gene="OJ1282_E10.6"
                     /note="contains full-length cDNA(s): AK107147"
                     /codon_start=1
                     /product="putative protein kinase"
                     /protein_id="BAD28451.1"
                     /db_xref="GI:50251413"
                     /translation="MKEGMSCFSWNPSLRCWHGSNDGGSPCSQAAQAASAGAGGAKKF
                     TLAQLSAATDGFHESNVVGEGGFGRVYRGRLEEGGQGLVAVKQLCHGGAQGTREFLVE
                     CMMLMMLHHPNLVSLVGYCADAGERLLVYEFLPRGSLDAHLFGRRPQEPPLALGWAAR
                     VRIAVGAARGLRYLHEVVTPPVIYRDLKASNILLDDDLNPRLSDFGLAKLGPVGDDTH
                     VSTRVMGTYGYCAPDYAMSGKLNVKSDVYSFGVVLLELITGRRAFDAASSDSESEDHQ
                     RFLLLRDWARPYLAGDRKRCFALADPALQGRYPRRAFYQLAVVASLCLRDNPNLRPSM

Alignment Scores:
Pred. No.:                2.11e+04        Length:          107284
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       15              Gaps:            0

US-09-858-332G-28 (1-6) x AP005290 (1-107284)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      54374 CATAATCATAATCATAAT 54391

RESULT 589
AP000641/c
LOCUS       AP000641              107607 bp    DNA     linear   PRI 15-MAR-2003
DEFINITION  Homo sapiens genomic DNA, chromosome 11q, clone:CMB9-44K17,
            complete sequence.
ACCESSION   AP000641
VERSION     AP000641.4  GI:15320454
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA
  JOURNAL   Published Only in Database (1999)
```

```
REFERENCE   2  (bases 1 to 107607)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-OCT-1999) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     On Aug 27, 2001 this sequence version replaced gi:11320560.
FEATURES             Location/Qualifiers
     source          1. .107607
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="CMB9-44K17"
ORIGIN

Alignment Scores:
Pred. No.:              2.12e+04        Length:         107607
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     8               Gaps:           0

US-09-858-332G-28 (1-6) x AP000641 (1-107607)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      90125 CATAACCACAACCATAAC 90108

RESULT 590
AP007600
LOCUS       AP007600              107855 bp    DNA     linear   HTG 28-DEC-2004
DEFINITION  Lotus corniculatus var. japonicus clone LjT45D24, *** SEQUENCING IN
            PROGRESS ***, 34 unordered pieces.
ACCESSION   AP007600
VERSION     AP007600.1  GI:56805913
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1
  AUTHORS   Kaneko,T., Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Structural Analysis of a Lotus japonicus Genome. XI. Sequence
            Features and Mapping of Nine hundred twenty-one TAC Clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 107855)
  AUTHORS   Sato,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-2004) Shusei Sato, Kazusa DNA Research Institute,
            Department of Plant Gene Research; 2-6-7 Kazusa-kamatari, Kisarazu,
            Chiba, 292-0818, Japan (E-mail:ssato@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/, Tel:81-438-52-3935(ex.2337),
            Fax:81-438-52-3934)
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 34 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1522: contig of 1522 bp in length
            *     1523     1622: gap of unknown length
            *     1623     2481: contig of 859 bp in length
            *     2482     2581: gap of unknown length
            *     2582     3688: contig of 1107 bp in length
            *     3689     3788: gap of unknown length
            *     3789     4731: contig of 943 bp in length
            *     4732     4831: gap of unknown length
            *     4832     5618: contig of 787 bp in length
            *     5619     5718: gap of unknown length
            *     5719     7093: contig of 1375 bp in length
            *     7094     7193: gap of unknown length
            *     7194     7972: contig of 779 bp in length
            *     7973     8072: gap of unknown length
            *     8073     9254: contig of 1182 bp in length
            *     9255     9354: gap of unknown length
            *     9355    11582: contig of 2228 bp in length
            *    11583    11682: gap of unknown length
            *    11683    12969: contig of 1287 bp in length
            *    12970    13069: gap of unknown length
            *    13070    14598: contig of 1529 bp in length
            *    14599    14698: gap of unknown length
            *    14699    16500: contig of 1802 bp in length
            *    16501    16600: gap of unknown length
            *    16601    18146: contig of 1546 bp in length
            *    18147    18246: gap of unknown length
            *    18247    20149: contig of 1903 bp in length
            *    20150    20249: gap of unknown length
            *    20250    22197: contig of 1948 bp in length
            *    22198    22297: gap of unknown length
            *    22298    24153: contig of 1856 bp in length
            *    24154    24253: gap of unknown length
            *    24254    25811: contig of 1558 bp in length
            *    25812    25911: gap of unknown length
            *    25912    28572: contig of 2661 bp in length
            *    28573    28672: gap of unknown length
            *    28673    31199: contig of 2527 bp in length
            *    31200    31299: gap of unknown length
            *    31300    34162: contig of 2863 bp in length
            *    34163    34262: gap of unknown length
            *    34263    36678: contig of 2416 bp in length
            *    36679    36778: gap of unknown length
            *    36779    39313: contig of 2535 bp in length
            *    39314    39413: gap of unknown length
            *    39414    42148: contig of 2735 bp in length
            *    42149    42248: gap of unknown length
            *    42249    46348: contig of 4100 bp in length
            *    46349    46448: gap of unknown length
            *    46449    49650: contig of 3202 bp in length
            *    49651    49750: gap of unknown length
            *    49751    53160: contig of 3410 bp in length
            *    53161    53260: gap of unknown length
            *    53261    60178: contig of 6918 bp in length
            *    60179    60278: gap of unknown length
            *    60279    65960: contig of 5682 bp in length
            *    65961    66060: gap of unknown length
            *    66061    71353: contig of 5293 bp in length
            *    71354    71453: gap of unknown length
            *    71454    76564: contig of 5111 bp in length
            *    76565    76664: gap of unknown length
            *    76665    84110: contig of 7446 bp in length
            *    84111    84210: gap of unknown length
            *    84211    90143: contig of 5933 bp in length
            *    90144    90243: gap of unknown length
            *    90244   102870: contig of 12627 bp in length
            *   102871   102970: gap of unknown length
            *   102971   107855: contig of 4885 bp in length.
FEATURES             Location/Qualifiers
     source          1. .107855
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="genomic DNA"
                     /variety="japonicus"
                     /db_xref="taxon:34305"
                     /clone="LjT45D24"
                     /clone_lib="LjT library"
                     /note="TAC clone:TM0801, synonym:Lotus japonicus"
     gap             1523. .1622
```

```
                /estimated_length=unknown
     gap        2482. .2581
                /estimated_length=unknown
     gap        3689. .3788
                /estimated_length=unknown
     gap        4732. .4831
                /estimated_length=unknown
     gap        5619. .5718
                /estimated_length=unknown
     gap        7094. .7193
                /estimated_length=unknown
     gap        7973. .8072
                /estimated_length=unknown
     gap        9255. .9354
                /estimated_length=unknown
     gap        11583. .11682
                /estimated_length=unknown
     gap        12970. .13069
                /estimated_length=unknown
     gap        14599. .14698
                /estimated_length=unknown
     gap        16501. .16600
                /estimated_length=unknown
     gap        18147. .18246
                /estimated_length=unknown
     gap        20150. .20249
                /estimated_length=unknown
     gap        22198. .22297
                /estimated_length=unknown
     gap        24154. .24253
                /estimated_length=unknown
     gap        25812. .25911
                /estimated_length=unknown
     gap        28573. .28672
                /estimated_length=unknown
     gap        31200. .31299
                /estimated_length=unknown
     gap        34163. .34262
                /estimated_length=unknown
     gap        36679. .36778
                /estimated_length=unknown
     gap        39314. .39413
                /estimated_length=unknown
     gap        42149. .42248
                /estimated_length=unknown
     gap        46349. .46448
                /estimated_length=unknown
     gap        49651. .49750
                /estimated_length=unknown
     gap        53161. .53260
                /estimated_length=unknown
     gap        60179. .60278
                /estimated_length=unknown
     gap        65961. .66060
                /estimated_length=unknown
     gap        71354. .71453
                /estimated_length=unknown
     gap        76565. .76664
                /estimated_length=unknown
     gap        84111. .84210
                /estimated_length=unknown
     gap        90144. .90243
                /estimated_length=unknown
     gap        102871. .102970
                /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              2.12e+04        Length:         107855
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0

DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AP007600 (1-107855)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      94204 CACAACCACAACCACAAC 94221

RESULT 591
AC103881/c
LOCUS       AC103881              108100 bp    DNA     linear   PRI 30-APR-2005
DEFINITION  Homo sapiens BAC clone RP11-638D14 from 2, complete sequence.
ACCESSION   AC103881
VERSION     AC103881.4  GI:19697537
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 108100)
  AUTHORS   Kozlowicz,A., Spalding,L. and Levy,A.
  TITLE     The sequence of Homo sapiens BAC clone RP11-638D14
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 108100)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-NOV-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3  (bases 1 to 108100)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAR-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   4  (bases 1 to 108100)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-APR-2005) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Mar 23, 2002 this sequence version replaced gi:18030160.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0638D14
            --------------.

            NOTICE:

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
```

```
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR:  pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-789A18, 2000 bp overlap;
            the clone sequenced to the right is RP11-515K12, 2000 bp overlap.
            Actual start of this clone is at base position 146042 of
            RP11-789A18; actual end is at base position 23686 of RP11-515K12.

            Polymorphisms have been identified between AC093113 and AC103881.

            A transposon was identified in a growth of the clone RP11-638D14,
            which is not a part of the submitted sequence.
FEATURES              Location/Qualifiers
     source           1. .108100
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /chromosome="2"
                      /clone="RP11-638D14"
                      /clone_lib="RPCI-11"
     misc_feature     77328. .77601
                      /note="CpG_island (%GC=62.4, o/e=0.86, #CpGs=24)"
ORIGIN

Alignment Scores:
Pred. No.:                2.13e+04        Length:         108100
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       8               Gaps:           0

US-09-858-332G-28 (1-6) x AC103881 (1-108100)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      17192 CACAATCATAATCACAAT 17175

RESULT 592
AC011663
LOCUS       AC011663               108582 bp    DNA     linear   PLN 19-JAN-2001
DEFINITION  Arabidopsis thaliana chromosome 1 BAC F5A18 genomic sequence,
            complete sequence.
ACCESSION   AC011663
VERSION     AC011663.7  GI:12324739
KEYWORDS    HTG.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 108582)
  AUTHORS   Lin,X., Kaul,S., Town,C.D., Benito,M.-I., Creasy,T.H., Haas,B.J.,
            Wu,D., Maiti,R., Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R.,
            Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.
  TITLE     Arabidopsis thaliana chromosome 1 BAC F5A18 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 108582)
  AUTHORS   Lin,X. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-1999) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
REFERENCE   3  (bases 1 to 108582)
  AUTHORS   Town,C.D. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
COMMENT     On Jan 19, 2001 this sequence version replaced gi:12280822.
            Address all correspondence to:at@tigr.org

            BAC clone F5A18 is from Arabidopsis thaliana chromosome 1
            The orientation of the sequence is from SP6 to T7 end of the BAC
            clone.
            Genes were identified by a combination of several methods: Gene
            prediction programs including Genscan+ (Chris Burge,
            http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHMM (Mark Borodovsky,
            http://genemark.biology.gatech.edu/GeneMark/), GlimmerA (a variant
            of GlimmerM, see Mihaela Pertea,
            http://www.tigr.org/softlab/glimmerm_htm/glimmerm.html, and
            GeneSplicer (Mihaela Pertea and Steven Salzberg, contact
            mpertea@tigr.org), searches of the complete sequence against a
            peptide database and the plant EST database at TIGR
            (http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to
            indicate the level of evidence for their annotation. Genes with
            similarity to other proteins are named after the database hits.
            Genes without significant peptide similarity but with EST
            similarity are named as  unknown  proteins. Genes without protein
            or EST similarity, that are predicted by more than two gene
            prediction programs over most of their length are annotated as
            hypothetical  proteins. Genes encoding tRNAs are predicted by
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
            Simple repeats are identified by repeatmasker (Arian Smit,
            http://ftp.genome.washington.edu/RM/RepeatMasker.html).
FEATURES              Location/Qualifiers
     source           1. .108582
                      /organism="Arabidopsis thaliana"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:3702"
                      /chromosome="1"
                      /clone="F5A18"
                      /ecotype="Columbia"
     gene             complement(1. .419)
                      /gene="F5A18.1"
                      /note="similar to N-term of hypothetical protein
                      GB:AAD55497 [Arabidopsis thaliana]"
     mRNA             complement(join(<1. .58,223. .>419))
                      /gene="F5A18.1"
     CDS              complement(join(<1. .58,223. .383))
                      /gene="F5A18.1"
                      /codon_start=1
                      /product="unknown protein, 3' partial; 383-1"
                      /protein_id="AAG52323.1"
                      /db_xref="GI:12324741"
                      /translation="MEELVGLLRIRVKRGINLAQRDTLSSDPFVVITMGSQVFSLSIS
                      VFDSNLIWSLENRLSILGCSYLACTLIEF"
     gene             complement(800. .1841)
                      /gene="F5A18.2"
                      /note="contains Pfam profile: PF00168 C2 domain; identical
                      to unknown protein GB:AAD55495 [Arabidopsis thaliana]"
     mRNA             complement(join(<800. .1230,1338. .1433,1696. .>1841))
                      /gene="F5A18.2"
     CDS              complement(join(940. .1230,1338. .1433,1696. .1833))
                      /gene="F5A18.2"
                      /codon_start=1
                      /product="unknown protein; 1833-940"
                      /protein_id="AAG52328.1"
                      /db_xref="GI:12324746"
                      /translation="MEKTEEEVEMKELVGLVRILVKRGIDLARRDALSSDPFVVITMG
                      PQKLKSFTVKNNCNPEWNEELTLAIEDPNEPVKLMVYDKDTFTADDKMGDAQIDMKPF
                      LDVHKLGLKELPHGKELKRIVPTRDNCLSEDSIIVSDNGKIVQDMILLLKNVECGKVE
                      IQLEWLKNPGGSGL"
     gene             complement(2463. .4165)
                      /gene="F5A18.3"
                      /note="contains Pfam profile: PF00168 C2 domain; identical
                      to unknown protein GB:AAD55494 [Arabidopsis thaliana]"
     mRNA             complement(join(<2463. .2506,2584. .2887,2970. .3065,
                      3753. .3868,4082. .4165))
                      /gene="F5A18.3"
     CDS              complement(join(2463. .2506,2584. .2887,2970. .3065,
                      3753. .3866))
                      /gene="F5A18.3"
                      /codon_start=1
                      /product="unknown protein; 3866-2463"
                      /protein_id="AAG52327.1"
```

```
                     /db_xref="GI:12324745"
                     /translation="MEDKPLGILRVHVKRGINLAIRDATTSDPYVVITLANQKLKTRV
                     INNNCNPVWNEQLTLSIKDVNDPIRLTVFDKDRFSGDDKMGDAEIDFRPFLEAHQMEL
                     DFQKLPNGCAIKRIRPGRTNCLAEESSITWSNGKIMQEMILRLKNVECGEVELMLEWT
                     DGPGCKGLGREGSKKTPWMPTKRLD"
repeat_region        complement(4891. .4917)
                     /rpt_family="(TA)n"
repeat_region        complement(7658. .7696)
                     /rpt_family="AT_rich"
repeat_region        complement(8038. .8067)
                     /rpt_family="AT_rich"
gene                 8232. .9267
                     /gene="F5A18.4"
                     /note="identical to unknown protein GB:AAD55493
                     [Arabidopsis thaliana]"
mRNA                 join(8232. .8849,8941. .9267)
                     /gene="F5A18.4"
CDS                  join(8615. .8849,8941. .9128)
                     /gene="F5A18.4"
                     /codon_start=1
                     /product="unknown protein; 8615-9128"
                     /protein_id="AAG52330.1"
                     /db_xref="GI:12324748"
                     /translation="MLLYNKQKKNQNAKGNRILISVTVLGSAGPIRFVAYEDDLVASV
                     IDTALKGYAREGRLPLLGSDFNDFLLYCPMVGPEALSTWDAIGSLGARNFMLCRKPEE
                     KKVEESNGRSDSTINGARKGGSLKAWINKSFNLKVSSH"
repeat_region        complement(12298. .12338)
                     /rpt_family="AT_rich"
repeat_region        complement(12474. .12518)
                     /rpt_family="AT_rich"
repeat_region        complement(12558. .12602)
                     /rpt_family="AT_rich"
repeat_region        complement(13031. .13052)
                     /rpt_family="AT_rich"
repeat_region        complement(13117. .13151)
                     /rpt_family="AT_rich"
gene                 13391. .16230
                     /gene="F5A18.5"
                     /note="identical to unknown protein GB:AAD55492
                     [Arabidopsis thaliana]"
mRNA                 join(13391. .13881,13982. .14119,14213. .14401,14500. .14637,
                     14738. .14856,14942. .15074,15159. .15288,15373. .15599,
                     15687. .16230)
                     /gene="F5A18.5"
CDS                  join(13405. .13881,13982. .14119,14213. .14401,14500. .14637,
                     14738. .14856,14942. .15074,15159. .15288,15373. .15599,
                     15687. .15968)
                     /gene="F5A18.5"
                     /codon_start=1
                     /product="unknown protein; 13405-15968"
                     /protein_id="AAG52333.1"
                     /db_xref="GI:12324751"
                     /translation="MDPIESVEYNGFETSNGNVDHGWKKVVYPKRHRKQKQADQATAN
                     GGKTVASNGTVANGGDNVFRSLEEQAEDRRRRILAAKMTAIDSDDDGVVRSKHRSNGY
                     GDDGYDFDGSDDEIATLKVEEVKKPKPKKEKKPKVSLPEAASKIDPLNLEAFLVEASE
                     SYASQPEIQLMRFADYFGRALSGVSSVQFPWVKMFKESPLSKLIEVPLAHIPEPVYKT
                     SVDWINHRPIEALGAFVLWAFDCILTDLAAQQGGAKGGKKGGQQTTSKSQVAIFVALA
                     MVLRRKPDALTNVLPTLRENPKYQGQDKLPVTVWMMAQASQGDIAVGLYSWAHNLLPV
                     VGNKNCNPQSRDLILQLVEKILTNPKARTILVNGAVRKGERLIPPPSFEILLRLTFPA
                     SSARVKATERFEAIYPLLKEVALAGAPGSKAMKQVTQQIFTFALKLAGEGNPVLAKEA
                     TAIAIWSVTQNFDCCKHWDNLYKENLEASVAVLKKLVEEWKEHSVKLSSSPNDALTLN
                     RTMKSFRLKNEEVITEGGANASLYKEADKSCKTISGRLSRGSGCLKGTAITLVFLAAA
                     GAAAAAVLSSNPEVTNELKNLVDSLELHEYINKYTEVISTALKK"
gene                 complement(16389. .17343)
                     /gene="F5A18.6"
                     /note="identical to residues 1 to 141 of unknown protein
                     GB:AAD55491 [Arabidopsis thaliana]"
mRNA                 complement(join(<16389. .16542,16628. .16722,16818. .16925,
                     17007. .17096,17195. .>17343))
                     /gene="F5A18.6"
CDS                  complement(join(16389. .16542,16628. .16722,16818. .16925,
                     17007. .17096,17195. .17323))
                     /gene="F5A18.6"
                     /codon_start=1
                     /product="unknown protein; 17323-16389"
                     /protein_id="AAG52336.1"
                     /db_xref="GI:12324754"
                     /translation="MSRCGSLGLYAPNALPSLSLKPRSVKSPFCITSHTKPNDTLLHN
                     VNKMRAKACDILGAKKTILAAQLGAVLATIDHPALAITGVNNQQELSSVVLDIGIISV
                     WYFLVMPPIIMNWLRVRWYRRKFFEMYLQFMFVFMFFPGLLLWAPFLNFRKFPRDPNM
                     KNPWDKPTDPDSIKNVYLKYPYATPEDYDLD"
repeat_region        complement(18481. .18504)
                     /rpt_family="AT_rich"
repeat_region        complement(20239. .20292)
                     /rpt_family="AT_rich"
repeat_region        complement(20821. .20871)
                     /rpt_family="AT_rich"
repeat_region        complement(21295. .21315)
                     /rpt_family="AT_rich"
gene                 21443. .24548
                     /gene="F5A18.7"
                     /note="similar to IFA-binding protein GB:AAD01202
                     [Arabidopsis thaliana]"
mRNA                 join(<21443. .22352,22429. .22563,22658. .23933,24037. .24548)
                     /gene="F5A18.7"
CDS                  join(21443. .22352,22429. .22563,22658. .23862)

Alignment Scores:
Pred. No.:              2.14e+04        Length:         108582
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     15              Gaps:           0

US-09-858-332G-28 (1-6) x AC011663 (1-108582)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      40959 CATAATCACAACCACAAC 40976

RESULT 593
AP007707
LOCUS       AP007707               108887 bp    DNA     linear   HTG 28-DEC-2004
DEFINITION  Lotus corniculatus var. japonicus chromosome 2 clone LjT06B24, ***
            SEQUENCING IN PROGRESS ***, 29 unordered pieces.
ACCESSION   AP007707
VERSION     AP007707.1  GI:56806017
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1
  AUTHORS   Kaneko,T., Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Structural Analysis of a Lotus japonicus Genome. XI. Sequence
            Features and Mapping of Nine hundred twenty-one TAC Clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 108887)
  AUTHORS   Sato,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-2004) Shusei Sato, Kazusa DNA Research Institute,
            Department of Plant Gene Research; 2-6-7 Kazusa-kamatari, Kisarazu,
            Chiba, 292-0818, Japan (E-mail:ssato@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/, Tel:81-438-52-3935(ex.2337),
            Fax:81-438-52-3934)
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 29 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
```

```
            *        1     1577: contig of 1577 bp in length
            *     1578     1677: gap of unknown length
            *     1678     2490: contig of 813 bp in length
            *     2491     2590: gap of unknown length
            *     2591     3526: contig of 936 bp in length
            *     3527     3626: gap of unknown length
            *     3627     4440: contig of 814 bp in length
            *     4441     4540: gap of unknown length
            *     4541     5899: contig of 1359 bp in length
            *     5900     5999: gap of unknown length
            *     6000     7901: contig of 1902 bp in length
            *     7902     8001: gap of unknown length
            *     8002     9251: contig of 1250 bp in length
            *     9252     9351: gap of unknown length
            *     9352    11106: contig of 1755 bp in length
            *    11107    11206: gap of unknown length
            *    11207    13024: contig of 1818 bp in length
            *    13025    13124: gap of unknown length
            *    13125    14917: contig of 1793 bp in length
            *    14918    15017: gap of unknown length
            *    15018    17611: contig of 2594 bp in length
            *    17612    17711: gap of unknown length
            *    17712    20343: contig of 2632 bp in length
            *    20344    20443: gap of unknown length
            *    20444    22428: contig of 1985 bp in length
            *    22429    22528: gap of unknown length
            *    22529    24945: contig of 2417 bp in length
            *    24946    25045: gap of unknown length
            *    25046    28468: contig of 3423 bp in length
            *    28469    28568: gap of unknown length
            *    28569    30889: contig of 2321 bp in length
            *    30890    30989: gap of unknown length
            *    30990    33418: contig of 2429 bp in length
            *    33419    33518: gap of unknown length
            *    33519    37220: contig of 3702 bp in length
            *    37221    37320: gap of unknown length
            *    37321    40892: contig of 3572 bp in length
            *    40893    40992: gap of unknown length
            *    40993    43872: contig of 2880 bp in length
            *    43873    43972: gap of unknown length
            *    43973    48566: contig of 4594 bp in length
            *    48567    48666: gap of unknown length
            *    48667    55015: contig of 6349 bp in length
            *    55016    55115: gap of unknown length
            *    55116    61245: contig of 6130 bp in length
            *    61246    61345: gap of unknown length
            *    61346    67764: contig of 6419 bp in length
            *    67765    67864: gap of unknown length
            *    67865    75533: contig of 7669 bp in length
            *    75534    75633: gap of unknown length
            *    75634    81322: contig of 5689 bp in length
            *    81323    81422: gap of unknown length
            *    81423    91591: contig of 10169 bp in length
            *    91592    91691: gap of unknown length
            *    91692   105350: contig of 13659 bp in length
            *   105351   105450: gap of unknown length
            *   105451   108887: contig of 3437 bp in length.
FEATURES             Location/Qualifiers
     source          1. .108887
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="genomic DNA"
                     /variety="japonicus"
                     /db_xref="taxon:34305"
                     /chromosome="2"
                     /clone="LjT06B24"
                     /clone_lib="LjT library"
                     /note="TAC clone:TM0961, synonym:Lotus japonicus"
     gap             1578. .1677
                     /estimated_length=unknown
     gap             2491. .2590
                     /estimated_length=unknown
     gap             3527. .3626
                     /estimated_length=unknown
     gap             4441. .4540
                     /estimated_length=unknown
     gap             5900. .5999
                     /estimated_length=unknown
     gap             7902. .8001
                     /estimated_length=unknown
     gap             9252. .9351
                     /estimated_length=unknown
     gap             11107. .11206
                     /estimated_length=unknown
     gap             13025. .13124
                     /estimated_length=unknown
     gap             14918. .15017
                     /estimated_length=unknown
     gap             17612. .17711
                     /estimated_length=unknown
     gap             20344. .20443
                     /estimated_length=unknown
     gap             22429. .22528
                     /estimated_length=unknown
     gap             24946. .25045
                     /estimated_length=unknown
     gap             28469. .28568
                     /estimated_length=unknown
     gap             30890. .30989
                     /estimated_length=unknown
     gap             33419. .33518
                     /estimated_length=unknown
     gap             37221. .37320
                     /estimated_length=unknown
     gap             40893. .40992
                     /estimated_length=unknown
     gap             43873. .43972
                     /estimated_length=unknown
     gap             48567. .48666
                     /estimated_length=unknown
     gap             55016. .55115
                     /estimated_length=unknown
     gap             61246. .61345
                     /estimated_length=unknown
     gap             67765. .67864
                     /estimated_length=unknown
     gap             75534. .75633
                     /estimated_length=unknown
     gap             81323. .81422
                     /estimated_length=unknown
     gap             91592. .91691
                     /estimated_length=unknown
     gap             105351. .105450
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:                2.14e+04        Length:          108887
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       14              Gaps:            0

US-09-858-332G-28 (1-6) x AP007707 (1-108887)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      58931 CACAACCACAACCATAAC 58948

RESULT 594
AC130804/c
LOCUS       AC130804               109715 bp    DNA     linear   HTG 02-FEB-2005
DEFINITION  Medicago truncatula clone mth2-1515, WORKING DRAFT SEQUENCE, 5
            ordered pieces.
ACCESSION   AC130804
VERSION     AC130804.44  GI:58418442
```

```
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 109715)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Medicago truncatula BAC Clone mth2-15l5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 109715)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-AUG-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 109715)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-FEB-2005) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Feb 2, 2005 this sequence version replaced gi:57900867.
            -------------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ----------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    26038: contig of 26038 bp in length
            *    26039    26138: gap of unknown length
            *    26139    44595: contig of 18457 bp in length
            *    44596    44695: gap of unknown length
            *    44696    59111: contig of 14416 bp in length
            *    59112    59211: gap of unknown length
            *    59212    81059: contig of 21848 bp in length
            *    81060    81159: gap of unknown length
            *    81160   109715: contig of 28556 bp in length.
FEATURES             Location/Qualifiers
     source          1. .109715
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3880"
                     /clone="mth2-1515"
                     /clone_lib="Medicago truncatula BAC library H2"
                     /note="This is one of two clones in the same well from
                     mth2-1515"
     gap             26039. .26138
                     /estimated_length=unknown
     gap             44596. .44695
                     /estimated_length=unknown
     gap             59112. .59211
                     /estimated_length=unknown
     gap             81060. .81159
                     /estimated_length=unknown
ORIGIN

Alignment Scores:
Pred. No.:              2.16e+04        Length:         109715
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     14              Gaps:           0

US-09-858-332G-28 (1-6) x AC130804 (1-109715)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      97417 CACAACCACAACCACAAT 97400

RESULT 595
AC126793/c
LOCUS       AC126793               109787 bp    DNA     linear   PLN 24-JUN-2004
DEFINITION  Medicago truncatula clone mth2-17m16, complete sequence.
ACCESSION   AC126793
VERSION     AC126793.13  GI:49170206
KEYWORDS    HTG.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 109787)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Medicago truncatula BAC Clone mth2-17m16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 109787)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 109787)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-SEP-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   4  (bases 1 to 109787)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   5  (bases 1 to 109787)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JAN-2004) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   6  (bases 1 to 109787)
  AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
            Cook,D., Kim,D. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUN-2004) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Jun 24, 2004 this sequence version replaced gi:41393273.
            -------------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ----------------------.
FEATURES             Location/Qualifiers
     source          1. .109787
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
```

/db_xref="taxon:3880"
/clone="mth2-17m16"
/clone_lib="Medicago truncatula BAC library H2"
ORIGIN

Alignment Scores:

Pred. No.:	2.16e+04	Length:	109787
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	15	Gaps:	0

US-09-858-332G-28 (1-6) x AC126793 (1-109787)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      55313 CATAATCATAATCATAAT 55296
```

RESULT 596
AE017225_22
WPCOMMENT
Sequence split into 53 fragments LOCUS AE017225 Accession AE017225

Fragment Name	Begin	End
AE017225_00	1	110000
AE017225_01	100001	210000
AE017225_02	200001	310000
AE017225_03	300001	410000
AE017225_04	400001	510000
AE017225_05	500001	610000
AE017225_06	600001	710000
AE017225_07	700001	810000
AE017225_08	800001	910000
AE017225_09	900001	1010000
AE017225_10	1000001	1110000
AE017225_11	1100001	1210000
AE017225_12	1200001	1310000
AE017225_13	1300001	1410000
AE017225_14	1400001	1510000
AE017225_15	1500001	1610000
AE017225_16	1600001	1710000
AE017225_17	1700001	1810000
AE017225_18	1800001	1910000
AE017225_19	1900001	2010000
AE017225_20	2000001	2110000
AE017225_21	2100001	2210000
AE017225_22	2200001	2310000
AE017225_23	2300001	2410000
AE017225_24	2400001	2510000
AE017225_25	2500001	2610000
AE017225_26	2600001	2710000
AE017225_27	2700001	2810000
AE017225_28	2800001	2910000
AE017225_29	2900001	3010000
AE017225_30	3000001	3110000
AE017225_31	3100001	3210000
AE017225_32	3200001	3310000
AE017225_33	3300001	3410000
AE017225_34	3400001	3510000
AE017225_35	3500001	3610000
AE017225_36	3600001	3710000
AE017225_37	3700001	3810000
AE017225_38	3800001	3910000
AE017225_39	3900001	4010000
AE017225_40	4000001	4110000
AE017225_41	4100001	4210000
AE017225_42	4200001	4310000
AE017225_43	4300001	4410000
AE017225_44	4400001	4510000
AE017225_45	4500001	4610000
AE017225_46	4600001	4710000
AE017225_47	4700001	4810000
AE017225_48	4800001	4910000
AE017225_49	4900001	5010000
AE017225_50	5000001	5110000
AE017225_51	5100001	5210000
AE017225_52	5200001	5228663

Continuation (23 of 53) of AE017225 from base 2200001 (AE017225 Bacillus anthracis str.

Alignment Scores:

Pred. No.:	2.16e+04	Length:	110000
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0

US-09-858-332G-28 (1-6) x AE017225_22 (1-110000)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      94741 CACAATCACAATCACAAT 94758
```

RESULT 597
AE017334_22
WPCOMMENT
Sequence split into 53 fragments LOCUS AE017334 Accession AE017334

Fragment Name	Begin	End
AE017334_00	1	110000
AE017334_01	100001	210000
AE017334_02	200001	310000
AE017334_03	300001	410000
AE017334_04	400001	510000
AE017334_05	500001	610000
AE017334_06	600001	710000
AE017334_07	700001	810000
AE017334_08	800001	910000
AE017334_09	900001	1010000
AE017334_10	1000001	1110000
AE017334_11	1100001	1210000
AE017334_12	1200001	1310000
AE017334_13	1300001	1410000
AE017334_14	1400001	1510000
AE017334_15	1500001	1610000
AE017334_16	1600001	1710000
AE017334_17	1700001	1810000
AE017334_18	1800001	1910000
AE017334_19	1900001	2010000
AE017334_20	2000001	2110000
AE017334_21	2100001	2210000
AE017334_22	2200001	2310000
AE017334_23	2300001	2410000
AE017334_24	2400001	2510000
AE017334_25	2500001	2610000
AE017334_26	2600001	2710000
AE017334_27	2700001	2810000
AE017334_28	2800001	2910000
AE017334_29	2900001	3010000
AE017334_30	3000001	3110000
AE017334_31	3100001	3210000
AE017334_32	3200001	3310000
AE017334_33	3300001	3410000
AE017334_34	3400001	3510000
AE017334_35	3500001	3610000
AE017334_36	3600001	3710000
AE017334_37	3700001	3810000
AE017334_38	3800001	3910000
AE017334_39	3900001	4010000
AE017334_40	4000001	4110000
AE017334_41	4100001	4210000
AE017334_42	4200001	4310000
AE017334_43	4300001	4410000
AE017334_44	4400001	4510000
AE017334_45	4500001	4610000
AE017334_46	4600001	4710000
AE017334_47	4700001	4810000

AE017334_48	4800001	4910000
AE017334_49	4900001	5010000
AE017334_50	5000001	5110000
AE017334_51	5100001	5210000
AE017334_52	5200001	5227419

Continuation (23 of 53) of AE017334 from base 2200001 (AE017334 Bacillus anthracis str.

Alignment Scores:

Pred. No.:	2.16e+04	Length:	110000
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0

US-09-858-332G-28 (1-6) x AE017334_22 (1-110000)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      94814 CACAATCACAATCACAAT 94831
```

RESULT 598
AP006628_6
WPCOMMENT
Sequence split into 9 fragments LOCUS AP006628 Accession AP006628

Fragment Name	Begin	End
AP006628_0	1	110000
AP006628_1	100001	210000
AP006628_2	200001	310000
AP006628_3	300001	410000
AP006628_4	400001	510000
AP006628_5	500001	610000
AP006628_6	600001	710000
AP006628_7	700001	810000
AP006628_8	800001	860631

Continuation (7 of 9) of AP006628 from base 600001 (AP006628 Onion yellows phytoplasma C

Alignment Scores:

Pred. No.:	2.16e+04	Length:	110000
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0

US-09-858-332G-28 (1-6) x AP006628_6 (1-110000)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      87738 CACAATCATAACCATAAC 87755
```

RESULT 599
CP000020_21/c
WPCOMMENT
Sequence split into 29 fragments LOCUS CP000020 Accession CP000020

Fragment Name	Begin	End
CP000020_00	1	110000
CP000020_01	100001	210000
CP000020_02	200001	310000
CP000020_03	300001	410000
CP000020_04	400001	510000
CP000020_05	500001	610000
CP000020_06	600001	710000
CP000020_07	700001	810000
CP000020_08	800001	910000
CP000020_09	900001	1010000
CP000020_10	1000001	1110000
CP000020_11	1100001	1210000
CP000020_12	1200001	1310000
CP000020_13	1300001	1410000
CP000020_14	1400001	1510000
CP000020_15	1500001	1610000
CP000020_16	1600001	1710000
CP000020_17	1700001	1810000
CP000020_18	1800001	1910000
CP000020_19	1900001	2010000
CP000020_20	2000001	2110000
CP000020_21	2100001	2210000
CP000020_22	2200001	2310000
CP000020_23	2300001	2410000
CP000020_24	2400001	2510000
CP000020_25	2500001	2610000
CP000020_26	2600001	2710000
CP000020_27	2700001	2810000
CP000020_28	2800001	2906179

Continuation (22 of 29) of CP000020 from base 2100001 (CP000020 Vibrio fischeri ES114 ch

Alignment Scores:

Pred. No.:	2.16e+04	Length:	110000
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0

US-09-858-332G-28 (1-6) x CP000020_21 (1-110000)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db      32649 CACAACCACAACCACAAC 32632
```

RESULT 600
CP000082_02/c
WPCOMMENT
Sequence split into 27 fragments LOCUS CP000082 Accession CP000082

Fragment Name	Begin	End
CP000082_00	1	110000
CP000082_01	100001	210000
CP000082_02	200001	310000
CP000082_03	300001	410000
CP000082_04	400001	510000
CP000082_05	500001	610000
CP000082_06	600001	710000
CP000082_07	700001	810000
CP000082_08	800001	910000
CP000082_09	900001	1010000
CP000082_10	1000001	1110000
CP000082_11	1100001	1210000
CP000082_12	1200001	1310000
CP000082_13	1300001	1410000
CP000082_14	1400001	1510000
CP000082_15	1500001	1610000
CP000082_16	1600001	1710000
CP000082_17	1700001	1810000
CP000082_18	1800001	1910000
CP000082_19	1900001	2010000
CP000082_20	2000001	2110000
CP000082_21	2100001	2210000
CP000082_22	2200001	2310000
CP000082_23	2300001	2410000
CP000082_24	2400001	2510000
CP000082_25	2500001	2610000
CP000082_26	2600001	2650701

Continuation (3 of 27) of CP000082 from base 200001 (CP000082 Psychrobacter arcticum 273

Alignment Scores:

Pred. No.:	2.16e+04	Length:	110000
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0

US-09-858-332G-28 (1-6) x CP000082_02 (1-110000)

```
Qy          1 HisAsnHisAsnHisAsn 6
```

```
                  ||||||||||||||||||
Db       72362 CATAATCATAATCATAAT 72345
```

Search completed: November 28, 2005, 13:40:39
Job time : 3992 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 28, 2005, 09:51:16 ; Search time 2606 Seconds
(without alignments)
107.722 Million cell updates/sec

Title: US-09-858-332G-28
Perfect score: 42
Sequence: 1 HNHNHN 6

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 1000 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09858332/runat_23112005_103529_19823/app_query.fasta_1.199
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=1000
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=600 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09858332_@CGN_1_1_5315_@runat_23112005_103529_19823 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	42	100.0	65	5	BQ587679	BQ587679 E012342-0
	2	42	100.0	92	7	CN926142	CN926142 000430AEP
	3	42	100.0	122	2	BF429042	BF429042 WHE1704_D
c	4	42	100.0	125	3	BJ340611	BJ340611 BJ340611
	5	42	100.0	154	3	BP061972	BP061972 BP061972
c	6	42	100.0	155	10	CE625685	CE625685 tigr-gss-
c	7	42	100.0	156	1	AW584439	AW584439 N210487e
	8	42	100.0	156	10	CE340511	CE340511 tigr-gss-
	9	42	100.0	157	7	CO278529	CO278529 EK130308.
	10	42	100.0	161	6	CB214559	CB214559 OML04839
	11	42	100.0	163	2	BF645838	BF645838 NF042B06E
c	12	42	100.0	163	3	BJ434601	BJ434601 BJ434601
	13	42	100.0	165	7	CV044762	CV044762 EST 6469
c	14	42	100.0	165	7	CV046698	CV046698 EST 8405
c	15	42	100.0	174	3	BJ429957	BJ429957 BJ429957
	16	42	100.0	174	7	CO269710	CO269710 EK095625.
	17	42	100.0	184	7	CV583709	CV583709 L_AH-aaa1
c	18	42	100.0	185	7	CV583842	CV583842 L_AH-aaa1
	19	42	100.0	193	2	BF202384	BF202384 WHE1764_A
	20	42	100.0	196	1	AI577627	AI577627 UI-R-AB0-
c	21	42	100.0	204	5	BQ855785	BQ855785 QGB27L15.
c	22	42	100.0	208	6	CB637138	CB637138 OSJNEa03F
	23	42	100.0	209	1	AU053723	AU053723 AU053723
c	24	42	100.0	209	5	BQ852441	BQ852441 QGB18A16.
	25	42	100.0	211	4	AK185713	AK185713 Mus muscu
	26	42	100.0	214	5	BQ586973	BQ586973 E012352-0
	27	42	100.0	216	1	AW158394	AW158394 za40h06.x
	28	42	100.0	220	6	CB214364	CB214364 OML04644
	29	42	100.0	222	8	R44827	R44827 yg31c02.s1
c	30	42	100.0	223	7	CO282440	CO282440 EK160504.
c	31	42	100.0	224	10	CE383475	CE383475 tigr-gss-
	32	42	100.0	228	7	CO484355	CO484355 GQ0204.B3
	33	42	100.0	231	1	AW689053	AW689053 NF014H03S
	34	42	100.0	233	1	AU073601	AU073601 AU073601
c	35	42	100.0	236	1	AU302349	AU302349 AU302349
c	36	42	100.0	237	7	CO023831	CO023831 EST788964
	37	42	100.0	240	1	AU073350	AU073350 AU073350
c	38	42	100.0	242	7	CO476909	CO476909 GQ0132.B3
c	39	42	100.0	245	1	AU303954	AU303954 AU303954
	40	42	100.0	247	5	BU994168	BU994168 HM05P14r
	41	42	100.0	250	6	CB606620	CB606620 AMGNNUC:C
	42	42	100.0	251	10	CW186051	CW186051 104_603_1
	43	42	100.0	253	9	CC779608	CC779608 ZMMBBc043
c	44	42	100.0	253	10	CG895908	CG895908 pastbac00
c	45	42	100.0	254	6	CB178412	CB178412 is38c09.y
	46	42	100.0	254	8	Z40263	Z40263 HSC1XE012 n
	47	42	100.0	258	3	BM522273	BM522273 ESSU0811
c	48	42	100.0	262	1	AJ497920	AJ497920 AJ497920
	49	42	100.0	262	10	CZ918340	CZ918340 4021009A0
	50	42	100.0	264	1	AU053106	AU053106 AU053106
	51	42	100.0	269	2	BF428793	BF428793 WHE1711_E
c	52	42	100.0	269	2	BE759303	BE759303 an_3656_A
c	53	42	100.0	271	10	CE679130	CE679130 tigr-gss-
	54	42	100.0	275	6	CB178653	CB178653 is38c09.x
	55	42	100.0	276	9	BH613103	BH613103 SALK_0337
	56	42	100.0	277	2	BG459579	BG459579 M77O22STM
	57	42	100.0	279	8	H46261	H46261 yo15d06.s1
c	58	42	100.0	281	6	CA264926	CA264926 SCAGRT304
	59	42	100.0	285	8	W40684	W40684 mc66a06.r1
c	60	42	100.0	286	2	BE759955	BE759955 an_2902_A
	61	42	100.0	287	1	AV419462	AV419462 AV419462
c	62	42	100.0	287	7	CO813747	CO813747 AGENCOURT
	63	42	100.0	287	10	CL392184	CL392184 ZMMBBb039
	64	42	100.0	289	7	CN252326	CN252326 EST018258
c	65	42	100.0	290	5	BQ986559	BQ986559 QGE9N08.y
	66	42	100.0	290	7	CF986668	CF986668 24212rsic
c	67	42	100.0	291	9	AZ278916	AZ278916 RPCI-23-1
	68	42	100.0	291	10	CL960835	CL960835 OsIFCC037
	69	42	100.0	292	1	AJ645412	AJ645412 AJ645412
c	70	42	100.0	295	7	CO301685	CO301685 EK181910.
c	71	42	100.0	297	5	BW876703	BW876703 BW876703
	72	42	100.0	298	2	BF547803	BF547803 UI-R-E0-b
	73	42	100.0	298	6	CD663439	CD663439 UCRHV18_0
	74	42	100.0	299	8	T10142	T10142 seq1163 b4H
c	75	42	100.0	299	9	BZ166889	BZ166889 CH230-377
	76	42	100.0	299	10	AJ592259	AJ592259 Arabidops
c	77	42	100.0	300	1	AJ473189	AJ473189 AJ473189
	78	42	100.0	305	1	AI485260	AI485260 EST243564
c	79	42	100.0	306	2	BE956148	BE956148 UI-M-BH4-
	80	42	100.0	307	2	BG046168	BG046168 saa49e08.

c	81	42	100.0	308	5	BY356096 BY356096 BY356096
c	82	42	100.0	310	9	AQ250124 AQ250124 T21D6-Sp6
	83	42	100.0	311	2	BF486624 BF486624 AT21076.5
c	84	42	100.0	313	1	AV560899 AV560899 AV560899
c	85	42	100.0	313	8	D43098 D43098 Rice
c	86	42	100.0	314	3	BJ351548 BJ351548 BJ351548
	87	42	100.0	315	1	AI483545 AI483545 EST249394
	88	42	100.0	315	7	CK116691 CK116691 B015P72 H
	89	42	100.0	316	8	H35592 H35592 EST110865 R
	90	42	100.0	317	1	AU227400 AU227400 AU227400
	91	42	100.0	317	8	D47535 D47535 RICS13103A
	92	42	100.0	317	8	R39818 R39818 yf51a04.s1
c	93	42	100.0	318	10	CE715092 CE715092 tigr-gss-
	94	42	100.0	319	1	AU061961 AU061961 AU061961
	95	42	100.0	320	7	CN927588 CN927588 000530AEP
c	96	42	100.0	321	5	BQ984974 BQ984974 QGE5J07.y
	97	42	100.0	321	7	CF985192 CF985192 24466rsic
	98	42	100.0	322	5	BU275786 BU275786 Cr_Emb_07
	99	42	100.0	322	10	CG685934 CG685934 ZMMBBC016
	100	42	100.0	324	1	AA264969 AA264969 LD08623.5
	101	42	100.0	324	1	AW664206 AW664206 hi06c06.x
c	102	42	100.0	324	8	DT050599 DT050599 COT_EB_E0
	103	42	100.0	325	3	BM536214 BM536214 EST589236
	104	42	100.0	325	7	CF962054 CF962054 9000rsice
c	105	42	100.0	327	3	BJ437953 BJ437953 BJ437953
	106	42	100.0	328	1	AW775956 AW775956 EST335021
	107	42	100.0	328	2	BE591060 BE591060 WHE0861_F
	108	42	100.0	329	10	BX002884 BX002884 Arabidops
	109	42	100.0	330	7	CV504055 CV504055 70570.1 M
	110	42	100.0	332	1	AV427149 AV427149 AV427149
c	111	42	100.0	333	8	DT051762 DT051762 COT_EP_H0
	112	42	100.0	334	2	BF640271 BF640271 NF027C05I
	113	42	100.0	334	2	BG139303 BG139303 EST479745
	114	42	100.0	334	2	BE192611 BE192611 EST331340
	115	42	100.0	334	5	BQ470935 BQ470935 HX04H16r
	116	42	100.0	335	2	BI234561 BI234561 RE30278.5
	117	42	100.0	336	1	AW127263 AW127263 M110204 G
	118	42	100.0	336	6	CB211509 CB211509 OML01789
	119	42	100.0	337	5	C93386 C93386 C93386 Dict
	120	42	100.0	337	6	CB212482 CB212482 OML02762
c	121	42	100.0	337	6	CF369337 CF369337 rg58e09.y
	122	42	100.0	338	2	BF293983 BF293983 WHE2162_D
	123	42	100.0	338	3	BM536213 BM536213 EST589235
c	124	42	100.0	338	5	BQ990783 BQ990783 QGF21B13.
c	125	42	100.0	338	6	CF369304 CF369304 rg58b07.y
	126	42	100.0	338	7	CN926847 CN926847 000502AEP
c	127	42	100.0	338	10	CG898233 CG898233 pastbac06
c	128	42	100.0	340	5	BX551988 BX551988 BX551988
	129	42	100.0	341	3	BP077382 BP077382 BP077382
	130	42	100.0	341	5	BY114663 BY114663 BY114663
	131	42	100.0	342	2	BI271692 BI271692 NF011B01F
	132	42	100.0	342	5	BQ854048 BQ854048 QGB22D09.
	133	42	100.0	343	1	AL373429 AL373429 MtBA57G07
	134	42	100.0	343	10	AG915691 AG915691 Drosophil
	135	42	100.0	344	5	BY104996 BY104996 BY104996
	136	42	100.0	345	5	BQ849201 BQ849201 QGA9D24.y
	137	42	100.0	346	1	AA186278 AA186278 T3492 MVA
	138	42	100.0	346	3	BP072036 BP072036 BP072036
	139	42	100.0	346	7	CO267388 CO267388 EK150015.
	140	42	100.0	347	1	AA098068 AA098068 mk18e01.r
	141	42	100.0	348	6	CF196076 CF196076 Azbi2.3_E
	142	42	100.0	349	5	BY776640 BY776640 BY776640
c	143	42	100.0	349	7	CV252723 CV252723 PX0015.B2
	144	42	100.0	352	7	CV546329 CV546329 SS2467.tx
	145	42	100.0	354	1	AW127340 AW127340 M110512 D
	146	42	100.0	354	2	BE604986 BE604986 WHE1713-1
c	147	42	100.0	355	2	BE240393 BE240393 EST404442
c	148	42	100.0	357	3	BI305914 BI305914 NL_2_G16
	149	42	100.0	357	3	BI419210 BI419210 LjNEST29f
c	150	42	100.0	357	6	CD852815 CD852815 DH0ALL9ZH
	151	42	100.0	357	8	W62873 W62873 md88a07.r1
	152	42	100.0	358	1	AU268115 AU268115 AU268115
	153	42	100.0	358	3	BM300074 BM300074 MCR054H05
c	154	42	100.0	358	5	BW840485 BW840485 BW840485
c	155	42	100.0	361	3	BJ340157 BJ340157 BJ340157
	156	42	100.0	362	1	AA409731 AA409731 EST01462
	157	42	100.0	362	7	CF990498 CF990498 22837rsic
	158	42	100.0	362	8	R40227 R40227 yf93f09.s1
	159	42	100.0	363	2	BE484332 BE484332 170900 BA
c	160	42	100.0	363	3	BI305814 BI305814 NL_1_N10
c	161	42	100.0	363	9	BH142412 BH142412 TDGDO50TH
c	162	42	100.0	364	3	BI306083 BI306083 NL_3_E04
c	163	42	100.0	364	3	BI306635 BI306635 NL_5_I10
c	164	42	100.0	365	3	BI306516 BI306516 NL_5_A01
	165	42	100.0	366	1	AI157489 AI157489 ue56e05.r
c	166	42	100.0	366	3	BP595103 BP595103 BP595103
	167	42	100.0	366	5	BY062997 BY062997 BY062997
	168	42	100.0	366	7	CN926563 CN926563 000501AEP
	169	42	100.0	366	8	R41647 R41647 yf88a10.s1
	170	42	100.0	367	5	BY301184 BY301184 BY301184
c	171	42	100.0	368	1	AJ761032 AJ761032 AJ761032
	172	42	100.0	368	5	BY083645 BY083645 BY083645
	173	42	100.0	368	7	CF963928 CF963928 13692rsic
	174	42	100.0	369	1	AU039300 AU039300 AU039300
c	175	42	100.0	369	3	BJ356615 BJ356615 BJ356615
c	176	42	100.0	369	3	BP630263 BP630263 BP630263
	177	42	100.0	370	2	BF292801 BF292801 WHE2165_A
	178	42	100.0	370	5	C93099 C93099 C93099 Dict
	179	42	100.0	371	2	BF643583 BF643583 NF092A09E
	180	42	100.0	371	2	BE521566 BE521566 M20D1STM
	181	42	100.0	371	3	BP623064 BP623064 BP623064
	182	42	100.0	372	7	CF991958 CF991958 17088rsic
c	183	42	100.0	373	1	AI016228 AI016228 ot83h12.s
c	184	42	100.0	373	2	BG662150 BG662150 Ljirnpest
	185	42	100.0	373	3	BI628630 BI628630 RH57130.5
	186	42	100.0	373	5	BY168107 BY168107 BY168107
	187	42	100.0	373	9	BZ545747 BZ545747 OGAFP26TC
	188	42	100.0	374	2	BI242911 BI242911 RE40419.5
c	189	42	100.0	374	3	BI306132 BI306132 NL_3_H06
	190	42	100.0	374	5	BY065744 BY065744 BY065744
c	191	42	100.0	374	9	BZ881398 BZ881398 CH240_218
	192	42	100.0	375	1	AI109054 AI109054 GH07561.5
c	193	42	100.0	375	3	BJ439631 BJ439631 BJ439631
c	194	42	100.0	375	3	BP803155 BP803155 BP803155
c	195	42	100.0	375	3	BQ120829 BQ120829 EST606405
	196	42	100.0	376	2	BF695474 BF695474 601851868
c	197	42	100.0	376	3	BP866389 BP866389 BP866389
	198	42	100.0	377	3	BP073663 BP073663 BP073663
	199	42	100.0	377	5	BQ982016 BQ982016 QGE15F05.
c	200	42	100.0	377	6	CF369358 CF369358 rg58g06.y
	201	42	100.0	378	5	BY079363 BY079363 BY079363
	202	42	100.0	379	5	BQ583906 BQ583906 E011887-0
	203	42	100.0	380	3	BP078874 BP078874 BP078874
	204	42	100.0	380	7	CN572644 CN572644 rg90a05.x
	205	42	100.0	381	7	CO472973 CO472973 GQ0045.B3
	206	42	100.0	381	8	W34128 W34128 mb81c05.r1
c	207	42	100.0	383	2	BG588299 BG588299 EST490108
	208	42	100.0	383	5	BQ805889 BQ805889 WHE3572_B
c	209	42	100.0	384	5	BW690941 BW690941 BW690941
	210	42	100.0	384	9	BH883598 BH883598 hw45f09.g
	211	42	100.0	385	6	CB826676 CB826676 LjNEST65f
	212	42	100.0	386	2	BF068492 BF068492 st82f08.y
	213	42	100.0	386	8	Z26004 Z26004 ATTS1281 Gr
	214	42	100.0	387	1	AI442701 AI442701 sa85c07.y
	215	42	100.0	387	8	T03482 T03482 IB36 Infant
	216	42	100.0	388	10	CW292826 CW292826 104_773_1
	217	42	100.0	389	1	AW912149 AW912149 uf42b08.y
c	218	42	100.0	390	1	AW332717 AW332717 S12A5 AGS
	219	42	100.0	390	5	BQ980435 BQ980435 QGE11A15.
	220	42	100.0	390	5	BY086496 BY086496 BY086496
	221	42	100.0	391	5	BY077809 BY077809 BY077809
	222	42	100.0	391	8	W96855 W96855 mf95e07.r1
c	223	42	100.0	391	9	CC153019 CC153019 CSU-K34.1
	224	42	100.0	392	2	BE637465 BE637465 WHE0859_E
c	225	42	100.0	392	5	BW862216 BW862216 BW862216
c	226	42	100.0	393	1	AU040894 AU040894 AU040894

	227	42	100.0	393	9	AZ254386 AZ254386 Gm_ISb001
	228	42	100.0	394	1	AJ770749 AJ770749
c	229	42	100.0	394	1	AU040954 AU040954 AU040954
c	230	42	100.0	394	1	AU040973 AU040973 AU040973
	231	42	100.0	395	1	AU040007 AU040007 AU040007
c	232	42	100.0	395	1	AU040972 AU040972 AU040972
c	233	42	100.0	395	3	BP588507 BP588507 BP588507
	234	42	100.0	395	6	CF196074 CF196074 Azbi2.3_E
	235	42	100.0	396	2	BG139519 BG139519 EST480057
c	236	42	100.0	396	5	BW876778 BW876778 BW876778
	237	42	100.0	396	7	CF959995 CF959995 8512rsice
	238	42	100.0	397	1	AI899397 AI899397 EST268840
	239	42	100.0	397	7	CF966741 CF966741 12584rsic
	240	42	100.0	398	1	AA004078 AA004078 mg81g02.r
	241	42	100.0	398	5	BY069921 BY069921 BY069921
	242	42	100.0	399	1	AV428528 AV428528 AV428528
	243	42	100.0	400	1	AL364520 AL364520 AL364520
	244	42	100.0	400	2	BF292873 BF292873 WHE2165_H
	245	42	100.0	400	2	BF293137 BF293137 WHE2164_A
	246	42	100.0	401	6	CF542925 CF542925 S014680-0
c	247	42	100.0	401	8	DR590716 DR590716 WS00829.B
	248	42	100.0	402	1	AI408694 AI408694 EST236985
	249	42	100.0	402	3	BI623556 BI623556 RH55856.5
	250	42	100.0	402	3	BM443064 BM443064 EBro02_SQ
	251	42	100.0	403	1	AI490526 AI490526 EST249060
	252	42	100.0	403	2	BG277971 BG277971 ux64h12.y
	253	42	100.0	403	9	AQ121290 AQ121290 HS_3075_A
	254	42	100.0	404	1	AA003697 AA003697 mg60g09.r
c	255	42	100.0	404	3	BP848836 BP848836 BP848836
	256	42	100.0	405	1	AI109877 AI109877 GH09260.5
	257	42	100.0	405	2	BF201722 BF201722 WHE1774_H
	258	42	100.0	405	2	BF324706 BF324706 su14c11.y
	259	42	100.0	405	6	CD157052 CD157052 ML1-0050U
	260	42	100.0	405	6	CD157053 CD157053 ML1-0050U
	261	42	100.0	405	6	CD157059 CD157059 ML1-0050U
	262	42	100.0	405	6	CD157120 CD157120 ML1-0050U
	263	42	100.0	405	6	CD157232 CD157232 ML1-0050U
	264	42	100.0	405	6	CD157273 CD157273 ML1-0050U
	265	42	100.0	405	6	CD157278 CD157278 ML1-0050U
	266	42	100.0	405	6	CF195899 CF195899 Azbi2.1_A
c	267	42	100.0	405	10	CW693980 CW693980 OG_BBa005
c	268	42	100.0	406	3	BJ428470 BJ428470 BJ428470
c	269	42	100.0	406	9	B23897 B23897 F17L16TR IG
c	270	42	100.0	408	1	AV811060 AV811060 AV811060
	271	42	100.0	408	3	BP028270 BP028270 BP028270
	272	42	100.0	408	5	BQ843997 BQ843997 QGA12E19.
	273	42	100.0	409	3	BP049169 BP049169 BP049169
	274	42	100.0	410	1	AA475682 AA475682 vh16b04.r
	275	42	100.0	410	3	BP029304 BP029304 BP029304
	276	42	100.0	411	5	BY288035 BY288035 BY288035
c	277	42	100.0	411	9	AQ010728 AQ010728 F27P2TFB
c	278	42	100.0	411	11	FR0042428 AL129920 Fugu rubr
	279	42	100.0	412	8	W81787 W81787 me92f05.r1
	280	42	100.0	412	11	FR0042427 AL129919 Fugu rubr
c	281	42	100.0	413	1	AV788333 AV788333 AV788333
	282	42	100.0	413	2	BG669342 BG669342 DRNAAED03
c	283	42	100.0	413	7	CN193652 CN193652 rg90a05.y
	284	42	100.0	413	9	AQ121286 AQ121286 HS_3075_A
	285	42	100.0	414	5	BQ864704 BQ864704 QGC27H01.
	286	42	100.0	415	10	BX002883 BX002883 Arabidops
	287	42	100.0	416	3	BM270929 BM270929 sak03c05.
	288	42	100.0	416	5	BY247437 BY247437 BY247437
	289	42	100.0	417	1	AU034405 AU034405 AU034405
	290	42	100.0	417	1	AW159517 AW159517 zb03d07.x
	291	42	100.0	417	2	BE606870 BE606870 WHE0902_A
	292	42	100.0	417	7	CN571761 CN571761 rg58h10.x
	293	42	100.0	417	8	W98585 W98585 mg09b04.r1
	294	42	100.0	419	6	CF096176 CF096176 QHN22F11.
	295	42	100.0	419	7	CV546384 CV546384 SS2534.tx
	296	42	100.0	420	1	AW985649 AW985649 uf84a09.y
	297	42	100.0	420	7	CF991917 CF991917 23707rsic
	298	42	100.0	420	8	Z34015 Z34015 ATTS3063 Pe
c	299	42	100.0	420	9	AQ655474 AQ655474 Sheared D
	300	42	100.0	421	1	AV406807 AV406807 AV406807
c	301	42	100.0	421	2	BF595803 BF595803 sv04g09.y
	302	42	100.0	421	5	BU809035 BU809035 UL52PH05
	303	42	100.0	422	1	AV768209 AV768209 AV768209
	304	42	100.0	422	2	BG262466 BG262466 WHE0936_C
c	305	42	100.0	422	7	CF979659 CF979659 rg79e02.y
	306	42	100.0	423	1	AV420924 AV420924 AV420924
c	307	42	100.0	423	7	CN194535 CN194535 rg95g07.y
c	308	42	100.0	424	1	AV809042 AV809042 AV809042
c	309	42	100.0	424	6	CF369371 CF369371 rg58h10.y
	310	42	100.0	424	9	AQ171001 AQ171001 HS_2217_A
	311	42	100.0	425	5	BQ576463 BQ576463 sap17e08.
c	312	42	100.0	425	7	CN193561 CN193561 rg88g02.y
c	313	42	100.0	426	1	AI008498 AI008498 EST202949
	314	42	100.0	426	3	BI653112 BI653112 603300651
	315	42	100.0	426	3	BI798061 BI798061 H096C06 E
	316	42	100.0	426	7	CO328401 CO328401 EP09137.3
	317	42	100.0	427	1	AU033706 AU033706 AU033706
	318	42	100.0	427	1	AU033893 AU033893 AU033893
	319	42	100.0	427	1	AV416079 AV416079 AV416079
c	320	42	100.0	427	1	AV596087 AV596087 AV596087
c	321	42	100.0	427	3	BP805483 BP805483 BP805483
	322	42	100.0	427	6	CF542950 CF542950 S014678-0
	323	42	100.0	427	7	CK760961 CK760961 nad03-11c
c	324	42	100.0	427	7	CN729641 CN729641 EX20LIB3_
	325	42	100.0	427	8	DR699194 DR699194 An_3235.A
	326	42	100.0	427	8	H06653 H06653 y182b11.s1
	327	42	100.0	428	2	BG604753 BG604753 WHE0941_G
	328	42	100.0	429	1	AV421073 AV421073
	329	42	100.0	429	5	BY034160 BY034160 BY034160
	330	42	100.0	430	1	AI152323 AI152323 ud79a02.r
	331	42	100.0	430	2	BE604994 BE604994 WHE1713-1
	332	42	100.0	430	2	BF012441 BF012441 ux56d02.y
c	333	42	100.0	430	5	BU017925 BU017925 QHE17D20.
	334	42	100.0	430	7	CF987483 CF987483 15498rsic
c	335	42	100.0	430	7	CO320011 CO320011 EK281720.
c	336	42	100.0	431	2	BG071356 BG071356 H3097H01-
	337	42	100.0	431	6	CA715331 CA715331 wdk3c.pk0
	338	42	100.0	431	7	CO294672 CO294672 EK207425.
	339	42	100.0	432	7	CV260941 CV260941 WS02015.B
	340	42	100.0	432	7	CV269305 CV269305 WS0207.B2
	341	42	100.0	432	7	CV475084 CV475084 23562.1 D
	342	42	100.0	433	1	AJ611217 AJ611217 AJ611217
	343	42	100.0	434	8	N59345 N59345 yz86f10.s1
	344	42	100.0	435	10	CL878923 CL878923 abf27b10.
c	345	42	100.0	436	3	BM369212 BM369212 EBem07_SQ
	346	42	100.0	437	2	BF941012 BF941012 hu63b06.x
	347	42	100.0	437	3	BJ234355 BJ234355 BJ234355
c	348	42	100.0	437	6	CB964579 CB964579 NL8_E02 D
	349	42	100.0	437	7	CF965114 CF965114 9492rsice
	350	42	100.0	437	7	CO336724 CO336724 EN13121.5
	351	42	100.0	438	1	AI488999 AI488999 EST247338
	352	42	100.0	438	2	BE758972 BE758972 an_3235 A
	353	42	100.0	438	7	CF986640 CF986640 23319rsic
	354	42	100.0	438	7	CK080515 CK080515 81076rsic
c	355	42	100.0	438	10	CZ590712 CZ590712 OA_BBa018
	356	42	100.0	440	1	AW596891 AW596891 sj84b04.y
	357	42	100.0	440	2	BE461297 BE461297 EST412716
	358	42	100.0	440	5	BQ804984 BQ804984 WHE3561_D
	359	42	100.0	441	8	H18998 H18998 yn51h01.s1
	360	42	100.0	442	3	BM577130 BM577130 170006871
	361	42	100.0	442	7	CF988556 CF988556 18328rsic
c	362	42	100.0	442	7	CN193991 CN193991 rg99g11.y
	363	42	100.0	444	2	BE604997 BE604997 WHE1713-1
	364	42	100.0	445	1	AA391089 AA391089 LD10018.5
c	365	42	100.0	446	1	AL379537 AL379537 MtBB45H08
	366	42	100.0	446	3	BJ324275 BJ324275 BJ324275
c	367	42	100.0	446	3	BJ434889 BJ434889 BJ434889
c	368	42	100.0	446	6	CA613496 CA613496 wr1.pk014
	369	42	100.0	447	1	AA636187 AA636187 vn15e07.r
	370	42	100.0	447	6	CA410761 CA410761 E031_F -P
	371	42	100.0	448	2	BE580088 BE580088 kq37a01.y
	372	42	100.0	448	7	CF965703 CF965703 14418rsic

c	373	42	100.0	448	9	AQ778216
	374	42	100.0	449	1	AI317046
	375	42	100.0	449	5	BY059656
	376	42	100.0	449	7	CF960519
	377	42	100.0	450	7	CF963362
	378	42	100.0	451	3	BQ244558
	379	42	100.0	451	5	BQ490899
	380	42	100.0	451	5	C94403
c	381	42	100.0	451	6	CB821385
	382	42	100.0	451	8	H11071
	383	42	100.0	452	1	AI721582
	384	42	100.0	452	1	AU269566
c	385	42	100.0	452	3	BP591415
c	386	42	100.0	452	8	DN373078
	387	42	100.0	453	1	AU040068
c	388	42	100.0	453	3	BP595105
c	389	42	100.0	453	7	CN193518
	390	42	100.0	454	7	CN242783
c	391	42	100.0	454	7	CN828903
	392	42	100.0	454	8	W50496
	393	42	100.0	455	1	AI183075
c	394	42	100.0	455	3	BJ432539
	395	42	100.0	456	1	AA444856
	396	42	100.0	456	2	BF483293
	397	42	100.0	456	7	CF985606
	398	42	100.0	457	2	BE654418
	399	42	100.0	457	5	C90662
	400	42	100.0	457	7	CF985050
	401	42	100.0	458	1	AA184544
	402	42	100.0	458	1	AL368151
	403	42	100.0	458	3	BQ246771
	404	42	100.0	458	6	CA667753
c	405	42	100.0	458	7	CO312414
	406	42	100.0	459	1	AU036977
	407	42	100.0	459	2	BE198421
	408	42	100.0	460	3	BM675713
	409	42	100.0	460	5	BQ492753
	410	42	100.0	460	6	CF197290
	411	42	100.0	460	9	BH017468
c	412	42	100.0	461	5	BQ853517
	413	42	100.0	461	5	BY249646
	414	42	100.0	461	9	BH144416
	415	42	100.0	462	3	BP926853
c	416	42	100.0	463	3	BM931692
c	417	42	100.0	463	3	BP043338
	418	42	100.0	463	7	CF992233
	419	42	100.0	464	2	BF293786
	420	42	100.0	464	2	BG088461
	421	42	100.0	464	2	BG140455
	422	42	100.0	464	5	BQ589046
	423	42	100.0	465	2	BG277124
c	424	42	100.0	465	3	BJ428195
c	425	42	100.0	465	3	BJ433147
	426	42	100.0	466	2	BG154056
	427	42	100.0	466	3	BQ246970
c	428	42	100.0	466	5	BQ982354
	429	42	100.0	466	6	CB714934
	430	42	100.0	466	7	CO344556
	431	42	100.0	466	7	CV056029
	432	42	100.0	466	7	CV057634
	433	42	100.0	466	7	CV060702
	434	42	100.0	466	7	CV061380
	435	42	100.0	466	8	W64428
	436	42	100.0	467	2	BI273062
	437	42	100.0	467	2	BB856288
c	438	42	100.0	467	3	BJ431306
c	439	42	100.0	467	5	BU545370
	440	42	100.0	467	7	CF974481
	441	42	100.0	467	7	CF974699
	442	42	100.0	467	7	CV057671
	443	42	100.0	467	9	AQ526634
	444	42	100.0	468	1	AA253897
	445	42	100.0	468	1	AW257025

AQ778216 HS_3206_A
AI317046 uk71g09.y
BY059656 BY059656
CF960519 12399rsic
CF963362 13642rsic
BQ244558 TaE15040H
BQ490899 EST00128
C94403 C94403 Dict
CB821385 EST 2236
H11071 ym09e08.s1
AI721582 fc29f06.x
AU269566 AU269566
BP591415 BP591415
DN373078 LIB3733-0
AU040068 AU040068
BP595105 BP595105
CN193518 rg88a10.y
CN242783 EST008655
CN828903 EL3189R B
W50496 md04g05.r1
AI183075 ub93e07.r
BJ432539 BJ432539
AA444856 vg57c09.r
BF483293 WHE1791_B
CF985606 24107rsic
BE654418 UI-M-AH1-
C90662 C90662 Dict
CF985050 17245rsic
AA184544 mt50h01.r
AL368151 MtBA22E07
BQ246771 TaE15005F
CA667753 wlsu1.pk0
CO312414 EK228707.
AU036977 AU036977
BE198421 ug78d07.y
BM675713 UI-E-EJ1-
BQ492753 EST01919
CF197290 Ir64vi2.1
BH017468 TDGCZ85TH
BQ853517 QGB20L16.
BY249646 BY249646
BH144416 TDGEK76TH
BP926853 BP926853
BM931692 UI-E-EJ1-
BP043338 BP043338
CF992233 16866rsic
BF293786 WHE2153_A
BG088461 H3153D07-
BG140455 EST480897
BQ589046 E012581-0
BG277124 ux63g03.y
BJ428195 BJ428195
BJ433147 BJ433147
BG154056 342 L1N01
BQ246970 TaE15002H
BQ982354 QGE16D15.
CB714934 AMGNNUC:M
CO344556 EP19507.3
CV056029 BNEL13B4
CV057634 BNEL2A7 B
CV060702 BNEL60b9
CV061380 BNEL68b7
W64428 me07a07.r1
BI273062 NF099B06F
BB856288 BB856288
BJ431306 BJ431306
BU545370 GM880004B
CF974481 PSU_np13\
CF974699 PSU_nupup
CV057671 BNEL2E1 B
AQ526634 HS_5193_A
AA253897 my31e10.r
AW257025 EST305162

	446	42	100.0	468	3	BI807757
c	447	42	100.0	468	6	CA647588
	448	42	100.0	468	6	CA715672
	449	42	100.0	468	8	CX303292
c	450	42	100.0	468	10	CW772270
c	451	42	100.0	469	6	CF370711
c	452	42	100.0	469	7	CN729649
	453	42	100.0	470	2	BE503307
c	454	42	100.0	471	1	AV422361
c	455	42	100.0	471	3	BJ380110
	456	42	100.0	471	5	BQ804588
	457	42	100.0	471	7	CF985613
c	458	42	100.0	472	3	BJ429835
c	459	42	100.0	472	3	BM891644
	460	42	100.0	473	6	CF192804
	461	42	100.0	473	8	W84975
c	462	42	100.0	473	9	BH357772
	463	42	100.0	474	3	BP888769
	464	42	100.0	475	7	CF962395
c	465	42	100.0	475	9	BZ621658
c	466	42	100.0	476	1	AU303780
c	467	42	100.0	476	3	BJ358156
c	468	42	100.0	476	3	BJ431249
c	469	42	100.0	476	3	BP529011
	470	42	100.0	476	7	CF961968
c	471	42	100.0	476	7	CV276803
	472	42	100.0	476	8	H09746
c	473	42	100.0	476	9	AQ659741
	474	42	100.0	477	1	AA762840
	475	42	100.0	477	3	BP749972
	476	42	100.0	477	3	BP878667
	477	42	100.0	477	3	BP879943
	478	42	100.0	478	6	CD910279
	479	42	100.0	479	1	AW224434
	480	42	100.0	479	3	BQ246847
	481	42	100.0	479	6	CB872086
	482	42	100.0	479	6	CF193670
c	483	42	100.0	479	11	PT013G05U
	484	42	100.0	480	1	AA667520
	485	42	100.0	480	1	AV423668
	486	42	100.0	480	2	BF293843
	487	42	100.0	480	3	BP080661
	488	42	100.0	480	7	CO271880
c	489	42	100.0	480	11	CR294373
	490	42	100.0	481	1	AI992710
c	491	42	100.0	481	3	BJ435882
	492	42	100.0	481	3	BP888020
	493	42	100.0	481	3	BP888943
	494	42	100.0	481	7	CF988072
	495	42	100.0	482	1	AJ776664
	496	42	100.0	482	1	AV422803
	497	42	100.0	482	6	CB881699
c	498	42	100.0	482	9	BH833860
c	499	42	100.0	483	1	AV537689
c	500	42	100.0	483	6	CB866638
	501	42	100.0	484	2	BF293358
	502	42	100.0	484	3	BJ369193
	503	42	100.0	484	3	BP036715
	504	42	100.0	484	5	BU010485
	505	42	100.0	485	1	AA949708
	506	42	100.0	486	1	AA259786
c	507	42	100.0	486	3	BJ386030
	508	42	100.0	486	7	CF961117
c	509	42	100.0	486	8	CX186077
c	510	42	100.0	487	10	CNS026FK
c	511	42	100.0	488	3	BJ441676
	512	42	100.0	488	9	AZ321550
	513	42	100.0	489	1	AA044548
c	514	42	100.0	489	5	BX865324
c	515	42	100.0	489	8	DR399696
	516	42	100.0	489	10	AG966239
c	517	42	100.0	491	1	AV557891
	518	42	100.0	491	3	BP043778

BI807757 B002D06 O
CA647588 wre1n.pk0
CA715672 wdk3c.pk0
CX303292 C08029F05
CW772270 OG_BBa008
CF370711 rg52g11.y
CN729649 EX20LIB3_
BE503307 7a15e04.x
AV422361 AV422361
BJ380110 BJ380110
BQ804588 WHE3556_E
CF985613 17031rsic
BJ429835 BJ429835
BM891644 sam42b09.
CF192804 15c21j2.r
W84975 mf42g01.r1
BH357772 CH230-95B
BP888769 BP888769
CF962395 9808siced
BZ621658 ig73e12.g
AU303780 AU303780
BJ358156 BJ358156
BJ431249 BJ431249
BP529011 BP529011
CF961968 13885rsic
CV276803 WS0141.B2
H09746 ym01b07.s1
AQ659741 Sheared D
AA762840 vs77g09.r
BP749972 BP749972
BP878667 BP878667
BP879943 BP879943
CD910279 G468.114O
AW224434 EST298321
BQ246847 TaE15004F
CB872086 HC06I16y
CF193670 16107j2.r
AL446113 Parameciu
AA667520 vv18a05.r
AV423668 AV423668
BF293843 WHE2153_F
BP080661 BP080661
CO271880 EK105028.
CR294373 mte1-10O1
AI992710 701493365
BJ435882 BJ435882
BP888020 BP888020
BP888943 BP888943
CF988072 16952rsic
AJ776664 AJ776664
AV422803 AV422803
CB881699 HM10J21w
BH833860 BACPP5-D0
AV537689 AV537689
CB866638 HC06I16w
BF293358 WHE2157_A
BJ369193 BJ369193
BP036715 BP036715
BU010485 QGJ13J11.
AA949708 LD29690.5
AA259786 va87c12.r
BJ386030 BJ386030
CF961117 11158rsic
CX186077 D02_45-26
AL183305 Tetraodon
BJ441676 BJ441676
AZ321550 1M0042I11
AA044548 mj12c02.r
BX865324 BX865324
DR399696 TKN105B10
AG966239 Drosophil
AV557891 AV557891
BP043778 BP043778

	519	42	100.0	491	7	CF963249 CF963249 12143rsic
	520	42	100.0	492	2	BF292932 BF292932 WHE2166_E
c	521	42	100.0	492	3	BJ358288 BJ358288 BJ358288
c	522	42	100.0	492	3	BJ435953 BJ435953 BJ435953
	523	42	100.0	492	6	CB868647 CB868647 HC08P11w
c	524	42	100.0	492	6	CB964758 CB964758 NL11_D05
	525	42	100.0	492	8	W46041 W46041 mc82h02.r1
	526	42	100.0	493	5	BU885165 BU885165 R021E09 P
	527	42	100.0	493	6	CB826688 CB826688 LjNEST65g
	528	42	100.0	494	3	BJ388347 BJ388347 BJ388347
c	529	42	100.0	494	7	CO310824 CO310824 EK222720.
	530	42	100.0	495	3	BM409788 BM409788 EST584115
	531	42	100.0	495	7	CF960382 CF960382 12469rsic
	532	42	100.0	496	5	C90140 C90140 Dict
	533	42	100.0	496	6	CD886243 CD886243 G118.101L
	534	42	100.0	497	7	CF988679 CF988679 17338rsic
	535	42	100.0	497	7	CO266833 CO266833 EK147719.
	536	42	100.0	497	9	AQ184606 AQ184606 HS_2212_A
	537	42	100.0	498	1	AW979496 AW979496 EST310517
c	538	42	100.0	498	5	BU014670 BU014670 QGJ7P15.y
	539	42	100.0	498	5	BX099249 BX099249 BX099249
	540	42	100.0	498	11	CR491970 CR491970 mth2-165J
	541	42	100.0	499	5	BQ489630 BQ489630 21-E9233-
c	542	42	100.0	499	5	BU991516 BU991516 HD07C14r
	543	42	100.0	499	7	CF991700 CF991700 16342rsic
	544	42	100.0	499	7	CV548502 CV548502 3032HAEME
c	545	42	100.0	499	9	CC175193 CC175193 ZMMBBc029
	546	42	100.0	500	1	AA110311 AA110311 mp51d09.r
	547	42	100.0	500	3	BJ365115 BJ365115 BJ365115
c	548	42	100.0	500	3	BP188517 BP188517 BP188517
c	549	42	100.0	500	5	BU991888 BU991888 HD08D17r
c	550	42	100.0	500	6	CB660142 CB660142 OSJNEd01K
	551	42	100.0	500	7	CK041708 CK041708 48024rsic
c	552	42	100.0	500	10	CL892616 CL892616 abg16c02.
	553	42	100.0	501	6	CD888217 CD888217 G118.107G
c	554	42	100.0	502	3	BJ435042 BJ435042 BJ435042
	555	42	100.0	502	7	CJ346215 CJ346215 CJ346215
c	556	42	100.0	502	10	CW260551 CW260551 104_727_1
c	557	42	100.0	502	10	CW339495 CW339495 104_840_1
	558	42	100.0	503	1	AW908751 AW908751 uf57a05.y
	559	42	100.0	503	3	BQ106494 BQ106494 fc1077.e
	560	42	100.0	503	9	BZ417164 BZ417164 if75g09.g
	561	42	100.0	504	2	BF293488 BF293488 WHE2157_D
	562	42	100.0	504	3	BJ365128 BJ365128 BJ365128
	563	42	100.0	504	3	BQ253175 BQ253175 san70d03.
	564	42	100.0	504	7	CK618125 CK618125 ml05h06.y
c	565	42	100.0	505	3	BM406192 BM406192 EST580519
	566	42	100.0	505	5	BY248500 BY248500 BY248500
	567	42	100.0	505	6	CA542334 CA542334 C0617F12-
	568	42	100.0	505	7	CF963876 CF963876 14637rsic
	569	42	100.0	506	1	AA049613 AA049613 mj36b06.r
	570	42	100.0	506	3	BJ365024 BJ365024 BJ365024
c	571	42	100.0	507	1	AV541062 AV541062 AV541062
c	572	42	100.0	507	2	BE785726 BE785726 601475292
c	573	42	100.0	507	3	BJ565074 BJ565074 BJ565074
c	574	42	100.0	507	5	BQ472467 BQ472467 HB09H21T
	575	42	100.0	508	1	AI490297 AI490297 EST248623
c	576	42	100.0	508	1	AL368573 AL368573 MtBA25C11
	577	42	100.0	508	6	CF605814 CF605814 RADIC01_0
	578	42	100.0	508	7	CF991731 CF991731 17944rsic
	579	42	100.0	509	1	AA636864 AA636864 vr20g04.r
	580	42	100.0	509	2	BF177637 Ljirnpest
	581	42	100.0	509	7	CF991173 CF991173 18042rsic
c	582	42	100.0	509	7	CO333305 CO333305 EK306440.
	583	42	100.0	509	8	CX535329 CX535329 s13dNF39G
c	584	42	100.0	510	6	CA890073 CA890073 B0156H06-
	585	42	100.0	511	2	BG060120 BG060120 naf49b10.
c	586	42	100.0	511	3	BJ344984 BJ344984 BJ344984
	587	42	100.0	511	5	BQ985398 BQ985398 QGE6L08.y
	588	42	100.0	511	6	CD340364 CD340364 StrPu536.
	589	42	100.0	511	10	CL283244 CL283244 ZMMBBb062
	590	42	100.0	511	11	CR297245 CR297245 mte1-14A4
	591	42	100.0	512	1	AW720113 AW720113 LjNEST14b
c	592	42	100.0	512	3	BJ384097 BJ384097 BJ384097
	593	42	100.0	512	6	CF083291 CF083291 QHL1A12.y
	594	42	100.0	512	7	CO188836 CO188836 EK042146.
c	595	42	100.0	513	2	BI094976 BI094976 sae02c07.
c	596	42	100.0	513	7	CN827064 CN827064 EL2007R B
c	597	42	100.0	514	1	AV567232 AV567232 AV567232
	598	42	100.0	514	3	BM886283 BM886283 sam14f07.
	599	42	100.0	514	7	CF965313 CF965313 11490rsic
c	600	42	100.0	514	7	CN827673 CN827673 EL2449R B
	601	42	100.0	514	10	CW395875 CW395875 fsbb001f0
c	602	42	100.0	515	3	BJ433449 BJ433449 BJ433449
	603	42	100.0	515	3	BJ458021 BJ458021 BJ458021
	604	42	100.0	515	5	BU780134 BU780134 SJEFUA07
c	605	42	100.0	516	1	AI600077 AI600077 EST251780
c	606	42	100.0	516	3	BI516316 BI516316 BB160022A
	607	42	100.0	516	3	BJ338204 BJ338204 BJ338204
	608	42	100.0	516	5	CA023404 CA023404 HZ46C09r
	609	42	100.0	516	7	CF985124 CF985124 20265rsic
	610	42	100.0	516	8	DN494093 DN494093 I077P28.5
	611	42	100.0	517	2	BE823305 BE823305 GM700020B
c	612	42	100.0	517	7	CV617028 CV617028 zcn27b02.
	613	42	100.0	517	9	BZ621657 BZ621657 ig73e12.b
c	614	42	100.0	517	10	CW386817 CW386817 fsbb001f0
	615	42	100.0	517	10	AG936697 AG936697 Drosophil
c	616	42	100.0	517	10	CG050799 CG050799 PUILQ63TD
	617	42	100.0	518	2	BE484331 BE484331 170899 BA
c	618	42	100.0	518	3	BP108193 BP108193 BP108193
c	619	42	100.0	518	5	BU673005 BU673005 NL_9_53 D
c	620	42	100.0	518	5	BU673252 BU673252 NL_13_3 D
	621	42	100.0	518	6	CA682687 CA682687 wlm96.pk0
c	622	42	100.0	518	6	CB965151 CB965151 NL17_E02
	623	42	100.0	519	2	BG362662 BG362662 sac07a11.
	624	42	100.0	519	3	BJ368463 BJ368463 BJ368463
	625	42	100.0	519	5	BW199285 BW199285 BW199285
	626	42	100.0	520	3	BJ234283 BJ234283 BJ234283
c	627	42	100.0	520	6	CA899924 CA899924 PCS00167M
	628	42	100.0	520	6	CF227970 CF227970 PtaXM0007
	629	42	100.0	521	1	AA263575 AA263575 LD06763.5
c	630	42	100.0	521	2	BE480786 BE480786 166019 BA
	631	42	100.0	521	3	BJ367435 BJ367435 BJ367435
	632	42	100.0	521	5	BW262603 BW262603 BW262603
	633	42	100.0	521	7	CF964678 CF964678 12289rsic
	634	42	100.0	521	9	AZ927641 AZ927641 476.dit27
c	635	42	100.0	521	11	CR356730 CR356730 Arabidops
	636	42	100.0	522	1	AA068855 AA068855 mm64d09.r
	637	42	100.0	522	7	CF986473 CF986473 19560rsic
	638	42	100.0	523	1	AU054056 AU054056 AU054056
	639	42	100.0	523	1	AA499431 AA499431 vi93d09.r
c	640	42	100.0	523	3	BJ445346 BJ445346 BJ445346
	641	42	100.0	525	3	BJ368647 BJ368647 BJ368647
c	642	42	100.0	525	3	BJ445787 BJ445787 BJ445787
	643	42	100.0	525	3	BQ245946 BQ245946 TaE15018B
	644	42	100.0	525	7	CF987770 CF987770 15134rsic
	645	42	100.0	525	7	CO867343 CO867343 Mddb5009b
c	646	42	100.0	525	9	BH552284 BH552284 BOHFR76TR
c	647	42	100.0	525	11	CR888025 CR888025 Sus scrof
c	648	42	100.0	526	5	BQ855876 BQ855876 QGB27P17.
	649	42	100.0	527	2	BF254321 BF254321 HVSMEf000
	650	42	100.0	527	2	BE431459 BE431459 EST336273
	651	42	100.0	527	7	CN252050 CN252050 EST017982
	652	42	100.0	527	10	CW397860 CW397860 fsbb001f0
	653	42	100.0	528	3	BI628871 BI628871 RH57633.5
	654	42	100.0	528	5	BQ504736 BQ504736 EST612139
c	655	42	100.0	528	6	CB264620 CB264620 50-E01465
	656	42	100.0	528	6	CD887215 CD887215 G118.104I
c	657	42	100.0	528	6	CF047982 CF047982 QCL16b01.
	658	42	100.0	528	7	CF963854 CF963854 12989rsic
c	659	42	100.0	528	7	CK862633 CK862633 33804 In
c	660	42	100.0	528	9	BH245304 BH245304 pSB0193 S
	661	42	100.0	529	1	AU052880 AU052880 AU052880
c	662	42	100.0	529	2	BG632329 BG632329 GH06583.3
	663	42	100.0	529	3	BJ324484 BJ324484 BJ324484
c	664	42	100.0	529	5	BQ848244 BQ848244 QGA6K19.y

	665	42	100.0	529	5	BW195240 BW195240 BW195240
c	666	42	100.0	529	6	CA604463 CA604463 wr1.pk003
	667	42	100.0	529	7	CF985622 CF985622 17719rsic
c	668	42	100.0	529	10	BX944821 BX944821 Arabidops
	669	42	100.0	530	1	AA009221 AA009221 mg95b09.r
c	670	42	100.0	530	1	AV563075 AV563075 AV563075
	671	42	100.0	530	6	CD891513 CD891513 G118.117I
c	672	42	100.0	530	10	CW337790 CW337790 104_837_1
	673	42	100.0	531	1	AU053519 AU053519 AU053519
c	674	42	100.0	531	3	BJ377384 BJ377384 BJ377384
c	675	42	100.0	531	6	CB965275 NL19_A06
c	676	42	100.0	531	6	CB967418 CB967418 NL52_B12
c	677	42	100.0	531	6	CB967427 CB967427 NL52_D11
	678	42	100.0	532	1	AA214909 AA214909 mu76f05.r
c	679	42	100.0	532	1	AL045178 AL045178 DKFZp434B
c	680	42	100.0	532	1	AV521205 AV521205 AV521205
	681	42	100.0	532	3	BI798956 BI798956 H121D04 E
	682	42	100.0	532	3	BJ613804 BJ613804 BJ613804
c	683	42	100.0	532	7	CK844809 CK844809 UI-R-AE0-
c	684	42	100.0	532	7	CN828846 CN828846 EL3153R B
	685	42	100.0	533	1	AA214902 AA214902 mu76d05.r
	686	42	100.0	533	1	AA606578 AA606578 vm82d02.r
	687	42	100.0	533	3	BJ367414 BJ367414 BJ367414
	688	42	100.0	533	3	BQ245912 BQ245912 TaE15018F
	689	42	100.0	533	8	W29675 W29675 mc07d02.r1
	690	42	100.0	533	9	AQ491972 AQ491972 V100B10 m
	691	42	100.0	534	1	AA259353 AA259353 va50g11.r
c	692	42	100.0	534	2	BI203727 BI203727 EST521767
c	693	42	100.0	534	2	BI204672 BI204672 EST522712
	694	42	100.0	534	3	BI575418 BI575418 RH31064.5
c	695	42	100.0	534	3	BM880042 BM880042 ku04c05.y
	696	42	100.0	534	3	BP041151 BP041151 BP041151
c	697	42	100.0	534	10	CZ481405 CZ481405 e03654-5p
	698	42	100.0	535	1	AV779286 AV779286 AV779286
c	699	42	100.0	535	2	BE980031 BE980031 UI-M-CG0-
	700	42	100.0	535	3	BI965338 BI965338 id37a05.y
	701	42	100.0	535	3	BJ422348 BJ422348 BJ422348
	702	42	100.0	535	7	CO209165 CO209165 WS00914.B
	703	42	100.0	536	1	AW657311 AW657311 110085 MA
	704	42	100.0	536	3	BQ128807 BQ128807 ii23c10.y
	705	42	100.0	536	5	BQ663225 BQ663225 HU01I06w
	706	42	100.0	536	7	CO330905 CO330905 EK296415.
	707	42	100.0	536	8	W98871 W98871 mf89d04.r1
	708	42	100.0	537	3	BJ334321 BJ334321 BJ334321
	709	42	100.0	537	7	CK082135 CK082135 80962rsic
	710	42	100.0	538	1	AI006575 AI006575 ue15e06.y
	711	42	100.0	538	3	BJ330406 BJ330406 BJ330406
	712	42	100.0	538	5	C22962 C22962 C22962 Dict
	713	42	100.0	539	2	BE432467 BE432467 EST398996
c	714	42	100.0	539	3	BJ427945 BJ427945 BJ427945
c	715	42	100.0	539	3	BJ443631 BJ443631 BJ443631
	716	42	100.0	539	5	BQ591141 BQ591141 S015011-0
	717	42	100.0	539	6	CD886897 CD886897 G118.103J
	718	42	100.0	539	7	CR929992 CR929992 CR929992
	719	42	100.0	539	8	CX023376 CX023376 Mdlv4-404
	720	42	100.0	539	9	AQ690487 AQ690487 nbxb0082K
	721	42	100.0	539	11	CR513351 CR513351 mth4-42I1
c	722	42	100.0	540	1	AA874794 AA874794 UI-R-E0-b
	723	42	100.0	540	1	AA466788 AA466788 vd90b09.r
	724	42	100.0	540	3	BJ365955 BJ365955 BJ365955
	725	42	100.0	540	5	BQ591744 BQ591744 E012617-0
	726	42	100.0	540	10	CZ308271 CZ308271 ZMMBF0004
c	727	42	100.0	541	3	BJ378627 BJ378627 BJ378627
c	728	42	100.0	541	3	BJ431589 BJ431589 BJ431589
	729	42	100.0	541	7	CF961429 CF961429 12099rsic
c	730	42	100.0	541	9	BZ295794 BZ295794 CG1681.f1
	731	42	100.0	542	1	AA004191 AA004191 mg82d06.r
	732	42	100.0	542	2	BE334866 BE334866 us90a11.y
	733	42	100.0	542	7	CF960953 CF960953 8110rsice
	734	42	100.0	542	7	CF966553 CF966553 10790rsic
	735	42	100.0	542	8	CX162697 CX162697 90653.1 M
	736	42	100.0	543	1	AA050176 AA050176 mj14b06.r
	737	42	100.0	543	1	AW736366 AW736366 EST332285
	738	42	100.0	543	2	BE690060 BE690060 uw63h01.y
	739	42	100.0	543	3	BJ369050 BJ369050 BJ369050
c	740	42	100.0	543	5	BU754653 BU754653 UI-1-BA0-
	741	42	100.0	543	5	BU873210 BU873210 Q052E03 P
	742	42	100.0	543	7	CF966531 CF966531 7915rsice
	743	42	100.0	543	7	CF991407 CF991407 18863rsic
c	744	42	100.0	544	3	BJ443661 BJ443661 BJ443661
	745	42	100.0	544	5	BQ995744 BQ995744 QGG10N15.
	746	42	100.0	544	6	CB967041 CB967041 NL47_A04
	747	42	100.0	544	7	CV231761 CV231761 W80195.B2
	748	42	100.0	545	5	BU002792 BU002792 QGG33A05.
c	749	42	100.0	545	9	AQ385558 AQ385558 RPCI11-14
	750	42	100.0	545	10	CW145025 CW145025 104_537_1
c	751	42	100.0	546	3	BJ429095 BJ429095 BJ429095
	752	42	100.0	546	5	BW198058 BW198058 BW198058
	753	42	100.0	546	7	CF965538 CF965538 7941rsice
c	754	42	100.0	546	10	CL971059 CL971059 OsIFCC021
	755	42	100.0	547	1	AW476461 AW476461 uq76c02.y
	756	42	100.0	547	6	CB881777 CB881777 HM10N14w
	757	42	100.0	547	7	CO428287 CO428287 UI-M-HW0-
	758	42	100.0	548	1	AV950170 AV950170 AV950170
	759	42	100.0	548	2	BE999015 BE999015 EST430738
	760	42	100.0	548	7	CF961410 CF961410 12411rsic
	761	42	100.0	548	8	CX768956 CX768956 UI-M-HV0-
c	762	42	100.0	548	11	TA300C04Q AL489501 T. brucei
	763	42	100.0	549	2	BF425834 BF425834 su16e06.y
	764	42	100.0	549	2	BE610676 BE610676 sq64e05.y
	765	42	100.0	549	3	BI566629 BI566629 RH36330.5
c	766	42	100.0	549	5	BX900515 BX900515 BX900515
	767	42	100.0	549	6	CB541711 CB541711 PVEPSE300
	768	42	100.0	549	6	CB864765 CB864765 HD07C14w
	769	42	100.0	549	6	CB865037 CB865037 HD08D17w
	770	42	100.0	549	6	CB865554 CB865554 HD09P09w
c	771	42	100.0	549	6	CB967111 CB967111 NL48_B05
	772	42	100.0	549	7	CF961941 CF961941 13585rsic
c	773	42	100.0	550	5	BU673593 BU673593 NL_17_56
c	774	42	100.0	550	5	BU673632 BU673632 NL_18_10
c	775	42	100.0	550	6	CB964701 CB964701 NL_10C08
c	776	42	100.0	550	6	CB966849 CB966849 NL43_F12
c	777	42	100.0	550	6	CD217896 CD217896 pgr1n.pk0
c	778	42	100.0	550	8	DT049364 DT049364 COT_DL_H0
c	779	42	100.0	550	8	DT052030 DT052030 COT_ET_B1
	780	42	100.0	550	9	BZ421435 BZ421435 if81a05.b
	781	42	100.0	551	1	AI484063 AI484063 EST249934
	782	42	100.0	551	3	BJ368806 BJ368806 BJ368806
	783	42	100.0	552	1	AA067755 AA067755 mm56a07.r
c	784	42	100.0	552	8	CX162378 CX162378 90314.1 M
c	785	42	100.0	553	1	AL368060 AL368060 MtBA21H11
	786	42	100.0	553	3	BJ234439 BJ234439 BJ234439
	787	42	100.0	554	3	BJ361240 BJ361240 BJ361240
	788	42	100.0	554	7	CN996627 CN996627 Mdfw2042a
	789	42	100.0	554	9	AZ145694 AZ145694 SP_0025_A
	790	42	100.0	555	3	BQ247194 BQ247194 TaE15028B
	791	42	100.0	555	11	DE040300 DE040300 Branchios
	792	42	100.0	556	1	AU052763 AU052763 AU052763
c	793	42	100.0	556	3	BJ355375 BJ355375 BJ355375
c	794	42	100.0	556	3	BJ430064 BJ430064 BJ430064
	795	42	100.0	556	3	BM068365 KS08022D1 BM068365
	796	42	100.0	556	3	BQ079763 BQ079763 san38a07.
	797	42	100.0	556	7	CF966272 CF966272 9007rsice
	798	42	100.0	557	1	AI771795 AI771795 EST252895
	799	42	100.0	557	1	AA215070 AA215070 mu76f04.r
	800	42	100.0	557	2	BF003639 BF003639 EST432137
	801	42	100.0	557	5	CA015954 CA015954 HV11K10u
	802	42	100.0	557	6	CB210072 CB210072 OML00352
	803	42	100.0	557	7	CO208032 CO208032 WS0094.B2
	804	42	100.0	558	3	BJ235055 BJ235055 BJ235055
	805	42	100.0	558	3	BQ246816 BQ246816 TaE15005A
	806	42	100.0	558	5	BU833613 BU833613 T050D02 P
	807	42	100.0	558	7	CF965219 CF965219 8104rsice
c	808	42	100.0	558	7	CV267250 CV267250 WS02031.B
	809	42	100.0	558	10	CW298288 CW298288 104_780_1
	810	42	100.0	558	10	CZ262749 CZ262749 OA_BBa01O

c 811	42	100.0	559	2	BG375827	BG375827 UI-R-CS0-
812	42	100.0	559	3	BJ339394	BJ339394 BJ339394
813	42	100.0	559	3	BJ365574	BJ365574 BJ365574
c 814	42	100.0	559	7	CV501340	CV501340 65844.1 M
815	42	100.0	560	1	AA870604	AA870604 vq23g08.r
816	42	100.0	560	1	AJ804555	AJ804555 AJ804555
817	42	100.0	560	1	AU265802	AU265802 AU265802
818	42	100.0	560	2	BF293514	BF293514 WHE2157_F
819	42	100.0	560	3	BJ329834	BJ329834 BJ329834
c 820	42	100.0	560	8	DR531448	DR531448 WS02735.C
c 821	42	100.0	561	1	AW940191	AW940191 GH05223.3
c 822	42	100.0	561	3	BJ384160	BJ384160 BJ384160
823	42	100.0	561	5	BQ741266	BQ741266 saq16d07.
c 824	42	100.0	562	1	AA899000	AA899000 NCP6H3T7
825	42	100.0	562	1	AU053309	AU053309 AU053309
826	42	100.0	562	1	AW033459	AW033459 EST277030
827	42	100.0	562	2	BI230005	BI230005 RE27921.5
828	42	100.0	562	5	BU494816	BU494816 PfESTOab6
829	42	100.0	562	5	BW266679	BW266679 BW266679
830	42	100.0	562	7	CF963651	CF963651 8463rsice
831	42	100.0	562	7	CO418287	CO418287 Mdfrt3033
832	42	100.0	562	8	DR118020	DR118020 RTMG1_10_
c 833	42	100.0	562	8	DR500740	DR500740 WS02921.C
c 834	42	100.0	562	9	BH142693	BH142693 TDGDR64TH
c 835	42	100.0	562	10	CG960494	CG960494 MBEET82TF
836	42	100.0	563	1	AI485460	AI485460 EST243781
837	42	100.0	563	5	BQ595357	BQ595357 S015251-0
838	42	100.0	563	7	CF966423	CF966423 9528rsice
839	42	100.0	564	1	AJ501488	AJ501488 AJ501488
840	42	100.0	564	1	AU266863	AU266863 AU266863
c 841	42	100.0	564	3	BM349273	BM349273 MEST247-A
842	42	100.0	564	7	CF961387	CF961387 13695rsic
843	42	100.0	564	7	CO185485	CO185485 EK009728.
c 844	42	100.0	564	7	CV498023	CV498023 62271.1 M
845	42	100.0	564	8	DR953422	DR953422 EST114496
846	42	100.0	565	1	AI543444	AI543444 SD10191.5
847	42	100.0	565	3	BJ464355	BJ464355 BJ464355
848	42	100.0	565	5	BW826927	BW826927 BW826927
849	42	100.0	565	7	CF959693	CF959693 7904rsice
850	42	100.0	565	7	CF965105	CF965105 13519rsic
851	42	100.0	565	8	DR513275	DR513275 WS02735.C
852	42	100.0	566	1	AU266866	AU266866 AU266866
c 853	42	100.0	566	3	BP772206	BP772206 BP772206
854	42	100.0	566	5	BU997619	BU997619 HI08I06r
855	42	100.0	566	7	CO330908	CO330908 EK296419.
856	42	100.0	566	7	CV474166	CV474166 22499.1 D
857	42	100.0	567	1	AI771834	AI771834 EST252934
858	42	100.0	567	3	BJ234323	BJ234323 BJ234323
859	42	100.0	567	6	CA486806	CA486806 WHE4336_B
860	42	100.0	567	10	CL565337	CL565337 OB__Ba003
c 861	42	100.0	568	5	BQ788682	BQ788682 WHE4152_F
c 862	42	100.0	568	10	CE687602	CE687602 tigr-gss-
c 863	42	100.0	569	10	CG289925	CG289925 OGWLU09TV
c 864	42	100.0	570	3	BI301924	BI301924 UI-R-DL0-
c 865	42	100.0	570	6	CD491431	CD491431 WHE3087_E
c 866	42	100.0	570	6	CD833765	CD833765 BN45.001H
867	42	100.0	571	2	BI167730	BI167730 RE08180.5
c 868	42	100.0	571	2	BI209560	BI209560 EST527600
c 869	42	100.0	571	3	BJ430176	BJ430176 BJ430176
c 870	42	100.0	571	3	BJ432285	BJ432285 BJ432285
c 871	42	100.0	571	5	BQ875885	BQ875885 QGI9G10.y
872	42	100.0	571	8	DR493768	DR493768 WS0293.BR
c 873	42	100.0	571	8	DR501290	DR501290 WS0293.B2
874	42	100.0	572	1	AI897787	AI897787 EST267230
875	42	100.0	572	3	BI614910	BI614910 RH44409.5
876	42	100.0	572	6	CF311500	CF311500 ABF--06-L
c 877	42	100.0	572	8	DR499565	DR499565 WS02918.C
878	42	100.0	573	1	AI455743	AI455743 LD35413.5
879	42	100.0	573	5	BQ452909	BQ452909 sao93a12.
880	42	100.0	573	7	CF990911	CF990911 19801rsic
881	42	100.0	573	7	CN884225	CN884225 010815AAS
c 882	42	100.0	573	9	BH740917	BH740917 gt25a09.g
c 883	42	100.0	573	9	CE146493	CE146493 tigr-gss-
884	42	100.0	573	10	CW670969	CW670969 OG_BBa002
885	42	100.0	574	3	BJ370337	BJ370337 BJ370337
886	42	100.0	574	5	BQ606564	BQ606564 BRY_2429
c 887	42	100.0	574	6	CB259311	CB259311 58-E9623-
888	42	100.0	574	6	CD900733	CD900733 G356.101J
c 889	42	100.0	574	7	CN249800	CN249800 EST015719
890	42	100.0	575	7	CF988960	CF988960 20647rsic
891	42	100.0	576	1	AI832576	AI832576 at70c06.x
c 892	42	100.0	576	3	BJ379607	BJ379607 BJ379607
c 893	42	100.0	576	3	BJ429672	BJ429672 BJ429672
894	42	100.0	576	7	CV493454	CV493454 37870.1 C
895	42	100.0	576	8	DR493160	DR493160 WS02921.C
896	42	100.0	577	1	AA268306	AA268306 va89c10.r
c 897	42	100.0	577	2	BF681983	BF681983 602116954
898	42	100.0	577	2	BI216979	BI216979 RE24484.5
c 899	42	100.0	577	3	BI504449	BI504449 BB170002B
c 900	42	100.0	577	3	BI507991	BI507991 BB170001A
c 901	42	100.0	577	5	BU886048	BU886048 R039F11 P
c 902	42	100.0	578	2	BI206299	BI206299 EST524339
903	42	100.0	578	2	BE195039	BE195039 HVSMEh008
904	42	100.0	578	6	CB723652	CB723652 UI-M-GK0-
905	42	100.0	578	7	CF988044	CF988044 24006rsic
906	42	100.0	578	10	AG937982	AG937982 Drosophil
907	42	100.0	579	1	AI489919	AI489919 EST248258
c 908	42	100.0	579	1	AW917174	AW917174 EST348478
c 909	42	100.0	579	3	BJ445045	BJ445045 BJ445045
910	42	100.0	579	5	BQ551283	BQ551283 H4008D04-
c 911	42	100.0	579	5	BU968429	BU968429 HB07H21r
912	42	100.0	579	5	BW193749	BW193749 BW193749
c 913	42	100.0	579	6	CB255238	CB255238 46-E01533
914	42	100.0	579	7	CF959561	CF959561 9814siced
c 915	42	100.0	580	3	BJ432758	BJ432758 BJ432758
916	42	100.0	580	8	DR490590	DR490590 WS02915.C
917	42	100.0	581	1	AI516758	AI516758 LD42796.5
918	42	100.0	581	3	BP956984	BP956984 BP956984
c 919	42	100.0	581	5	BQ753231	BQ753231 WHE4124_E
c 920	42	100.0	581	7	CK034523	CK034523 39533rsic
c 921	42	100.0	581	7	CK660246	CK660246 LP18649.5
c 922	42	100.0	581	8	CX250503	CX250503 1302747 N
923	42	100.0	582	3	BJ335809	BJ335809 BJ335809
c 924	42	100.0	582	3	BJ349932	BJ349932 BJ349932
925	42	100.0	582	3	BJ366050	BJ366050 BJ366050
926	42	100.0	582	3	BJ369186	BJ369186 BJ369186
927	42	100.0	583	1	AU263254	AU263254 AU263254
928	42	100.0	583	6	CB586337	CB586337 AMGNNUC:N
929	42	100.0	583	7	CF984909	CF984909 18592sice
930	42	100.0	583	7	CF991926	CF991926 16250rsic
931	42	100.0	584	1	AI485175	AI485175 EST243479
932	42	100.0	584	3	BI671164	BI671164 fp61g09.x
933	42	100.0	584	5	BU006125	BU006125 QGH10A07.
934	42	100.0	584	7	CK082860	CK082860 85761rsic
935	42	100.0	585	2	BE334877	BE334877 us90b10.y
936	42	100.0	585	2	BE402590	BE402590 CSB009E08
937	42	100.0	585	3	BI621864	BI621864 RH53285.5
938	42	100.0	585	3	BJ331129	BJ331129 BJ331129
939	42	100.0	585	3	BJ331788	BJ331788 BJ331788
940	42	100.0	585	3	BJ414630	BJ414630 BJ414630
941	42	100.0	585	5	BQ608170	BQ608170 BRY_4072
942	42	100.0	585	5	BQ609411	BQ609411 BRY_5340
c 943	42	100.0	585	10	AG945847	AG945847 Drosophil
944	42	100.0	586	1	AI771755	AI771755 EST252855
c 945	42	100.0	586	1	AJ273983	AJ273983 AJ273983
c 946	42	100.0	586	10	CG006418	CG006418 ZUADJ03TV
947	42	100.0	587	1	AV765898	AV765898 AV765898
948	42	100.0	587	6	CF638492	CF638492 D03_C02 F
949	42	100.0	587	8	CV850474	CV850474 zaa73d11.
950	42	100.0	588	3	BJ392720	BJ392720 BJ392720
c 951	42	100.0	588	3	BJ431964	BJ431964 BJ431964
c 952	42	100.0	588	5	BQ517355	BQ517355 EST624770
953	42	100.0	588	7	CF967065	CF967065 9827siced
954	42	100.0	588	8	DR491981	DR491981 WS02918.C
955	42	100.0	589	3	BJ333284	BJ333284 BJ333284
956	42	100.0	589	3	BJ369720	BJ369720 BJ369720

c 957	42	100.0	589	5	BQ740898	BQ740898 sap88g02.
c 958	42	100.0	589	6	CD878548	CD878548 AZO4.103A
959	42	100.0	589	7	CV748045	CV748045 SJL_35_00
c 960	42	100.0	589	10	CE626612	CE626612 tigr-gss-
961	42	100.0	590	3	BI636340	BI636340 SD17985.5
962	42	100.0	590	3	BJ335974	BJ335974 BJ335974
963	42	100.0	590	3	BJ367600	BJ367600 BJ367600
964	42	100.0	590	6	CF397969	CF397969 RTDS3_23_
c 965	42	100.0	590	8	DR499767	DR499767 WS02919.C
966	42	100.0	590	8	DR504163	DR504163 WS02710.C
c 967	42	100.0	590	8	DR521822	DR521822 WS02710.C
968	42	100.0	590	8	W64711	W64711 md71g01.r1
969	42	100.0	591	5	BW194484	BW194484 BW194484
970	42	100.0	592	1	AA033182	AA033182 mi36a03.r
971	42	100.0	592	1	AU060834	AU060834 AU060834
972	42	100.0	592	3	BI605883	BI605883 RH71886.5
c 973	42	100.0	592	3	BJ343184	BJ343184 BJ343184
974	42	100.0	592	3	BJ368828	BJ368828 BJ368828
c 975	42	100.0	592	3	BJ431986	BJ431986 BJ431986
c 976	42	100.0	592	3	BP796399	BP796399 BP796399
c 977	42	100.0	592	3	BQ241515	BQ241515 TaE05004B
978	42	100.0	592	7	CK277233	CK277233 EST723311
c 979	42	100.0	592	7	CN729680	CN729680 EX20LIB3_
980	42	100.0	592	8	W59446	W59446 md80d01.r1
981	42	100.0	593	1	AU284394	AU284394 AU284394
c 982	42	100.0	593	1	AW267860	AW267860 EST306138
983	42	100.0	593	1	AW458899	AW458899 sh16c10.y
984	42	100.0	593	2	BF292833	BF292833 WHE2165_D
985	42	100.0	593	3	BJ325409	BJ325409 BJ325409
986	42	100.0	593	6	CA835626	CA835626 MCS039B10
987	42	100.0	593	7	CF953321	CF953321 6565rsice
c 988	42	100.0	593	7	CN830103	CN830103 EL3936R B
989	42	100.0	593	7	CV496426	CV496426 74230.1 C
990	42	100.0	593	9	AZ452610	AZ452610 1M0252L04
c 991	42	100.0	594	2	BG644727	BG644727 EST506346
992	42	100.0	594	3	BI309203	BI309203 EST530613
c 993	42	100.0	594	3	BJ429400	BJ429400 BJ429400
c 994	42	100.0	594	3	BJ432574	BJ432574 BJ432574
995	42	100.0	594	7	CK079978	CK079978 79610rsic
996	42	100.0	595	3	BQ246854	BQ246854 TaE15004E
c 997	42	100.0	595	5	CA019330	CA019330 HV11K10r
998	42	100.0	595	7	CN781975	CN781975 EST00070
999	42	100.0	596	5	BU065402	BU065402 Fgr_7_H12
c1000	42	100.0	596	6	CB006593	CB006593 VVC035F02

ALIGNMENTS

RESULT 1
BQ587679
LOCUS BQ587679 65 bp mRNA linear EST 06-DEC-2002
DEFINITION E012342-024-010-P03-SP6 MPIZ-ADIS-024-leaf Beta vulgaris cDNA clone 024-010-P03 5-PRIME, mRNA sequence.
ACCESSION BQ587679
VERSION BQ587679.1 GI:26117261
KEYWORDS EST.
SOURCE Beta vulgaris
ORGANISM Beta vulgaris
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; Caryophyllales; Amaranthaceae; Beta.
REFERENCE 1 (bases 1 to 65)
AUTHORS Herwig,R., Schulz,B., Weisshaar,B., Hennig,S., Steinfath,M., Drungowski,M., Stahl,D., Wruck,W., Menze,A., O'Brien,J., Lehrach,H. and Radelof,U.
TITLE Construction of a 'unigene' cDNA clone set by oligonucleotide fingerprinting allows access to 25 000 potential sugar beet genes
JOURNAL Plant J. 32 (5), 845-857 (2002)
PUBMED 12472698
COMMENT Contact: Weisshaar B
ADIS DNA core facility at MPIZ
Max-Planck-Institute for Plant Breeding Research
Carl-von-Linne Weg 10, 50829 Koeln, Germany
Fax: 00492215062851
Email: weisshaa@mpiz-koeln.mpg.de
Insert Length: 65 Std Error: 0.00
Plate: 10 row: P column: 03
Seq primer: SP6; CATACGATTTAGGTGACACTATAG.
FEATURES Location/Qualifiers
source 1. .65
/organism="Beta vulgaris"
/mol_type="mRNA"
/cultivar="KWS2320 (double haploid, monogerm breeding line)"
/db_xref="GABI:185182"
/db_xref="taxon:161934"
/clone="024-010-P03"
/tissue_type="leaf"
/lab_host="EMDH10B"
/clone_lib="MPIZ-ADIS-024-leaf"
/note="Vector: pCMVSPORT6; Site_1: SalI; Site_2: NotI; cDNA library from sugar beet, library provided by KWS Kleinwanzlebener Saatzucht AG Einbeck, Germany, contact: b.schulz@kws.de; cloning sites SalI-NotI, primer sites and orientation: SP6-SalI-CCACGCGTCCG-5prime-cDNA-polyA-CC-NotI-T7; Note: Sequencing granted in the context of the GABI-Beet project, local PI: Dr. Katharina Schneider, coordinator: Prof. Christian Jung; Sequence submission managed by RZPD/GABI-Primary database:http://gabi.rzpd.de"
ORIGIN

Alignment Scores:

Pred. No.:	330	Length:	65
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	5	Gaps:	0

US-09-858-332G-28 (1-6) x BQ587679 (1-65)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         34 CACAACCACAACCATAAC 51
```

RESULT 2
CN926142
LOCUS CN926142 92 bp mRNA linear EST 07-JUN-2004
DEFINITION 000430AEPA001249HT (AEPA) Pinkie expanding leaf Malus x domestica cDNA clone AEPA001249, mRNA sequence.
ACCESSION CN926142
VERSION CN926142.1 GI:48398955
KEYWORDS EST.
SOURCE Malus x domestica
ORGANISM Malus x domestica
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids I; Rosales; Rosaceae; Maloideae; Malus.
REFERENCE 1 (bases 1 to 92)
AUTHORS Beuning,L., Bowen,J., Crowhurst,R., Gleave,A., Janssen,B., McArtney,S., Newcomb,R., Ross,G., Snowden,K., Walton,E. and Yauk,Y.
TITLE HortResearch Apple EST Project
JOURNAL Unpublished (2004)
COMMENT Contact: Gleave,A.
Sequencing Facility
The Horticulture and Food Research Institute of New Zealand Ltd
120 Mt Albert Rd, Mt Albert, Auckland, New Zealand
Tel: 00 64 09 815 4200
Fax: 00 64 09 815 4201
Email: est@hortresearch.co.nz.
FEATURES Location/Qualifiers
source 1. .92
/organism="Malus x domestica"
/mol_type="mRNA"
/db_xref="taxon:3750"

/clone="AEPA001249"
/tissue_type="Leaf"
/dev_stage="Expanding"
/clone_lib="(AEPA) Pinkie expanding leaf"
/note="Vector: pBK-CMV; Library sequenced by Genesis Research & Development"

ORIGIN

Alignment Scores:

Pred. No.:	478	Length:	92
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	7	Gaps:	0

US-09-858-332G-28 (1-6) x CN926142 (1-92)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         72 CACAACCACAACCACAAC 89
```

RESULT 3
BF429042
LOCUS BF429042 122 bp mRNA linear EST 29-NOV-2000
DEFINITION WHE1704_D08_H16ZS Wheat heat stressed spike cDNA library Triticum aestivum cDNA clone WHE1704_D08_H16, mRNA sequence.
ACCESSION BF429042
VERSION BF429042.1 GI:11440961
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 122)
AUTHORS Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Malatrasi,M., Miller,R., Nguyen,H.T., Rausch,C.J., Seaton,C.L., Tong,J.C. and Zhang,D.
TITLE The structure and function of the expressed portion of the wheat genomes - Heat stressed spike cDNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oandersn@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
Seq primer: Stratagene SK primer.
FEATURES Location/Qualifiers
source 1. .122
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE1704_D08_H16"
/tissue_type="Whole spike"
/dev_stage="Spikes at 5, 10, 15 and 20 days after anthesis"
/lab_host="E. coli SOLR"
/clone_lib="Wheat heat stressed spike cDNA library"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site_1: EcoRI; Site_2: XhoI; Spikes at 5, 10, 15 and 20 days after anthesis were heat stressed under two conditions at Texas Tech University (D. Zhang in HT Nguyrn lab): (1) at 38 C for 4 hours and (2) 5 days of cyclic treatment of 38 C for 4 hours. Total RNA and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give pBluescript phagemids in the TJ Close lab (Choi, Close, Fenton, Malatrasi) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."
ORIGIN

Alignment Scores:

Pred. No.:	645	Length:	122
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-28 (1-6) x BF429042 (1-122)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db          1 CACAACCACAACCACAAT 18
```

RESULT 4
BJ340611/c
LOCUS BJ340611 125 bp mRNA linear EST 07-MAR-2002
DEFINITION BJ340611 Dictyostelium discoideum cDNA library, AF Dictyostelium discoideum cDNA clone dda3c02 3', mRNA sequence.
ACCESSION BJ340611
VERSION BJ340611.1 GI:19248973
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 125)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the aggregation stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES Location/Qualifiers
source 1. .125
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dda3c02"
/sex="mat A"
/dev_stage="Aggregation stage"
/clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN

Alignment Scores:

Pred. No.:	662	Length:	125
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	3	Gaps:	0

US-09-858-332G-28 (1-6) x BJ340611 (1-125)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        121 CACAATCACAATCACAAT 104
```

RESULT 5
BP061972
LOCUS BP061972 154 bp mRNA linear EST 25-AUG-2004
DEFINITION BP061972 Lotus corniculatus var. japonicus mature nodule Lotus corniculatus var. japonicus cDNA clone GENf085g08 3', mRNA

```
            sequence.
ACCESSION   BP061972
VERSION     BP061972.1  GI:45618720
KEYWORDS    EST.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1  (bases 1 to 154)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Characteristics of the Lotus japonicus gene repertoire deduced from
            large-scale expressed sequence tag (EST) analysis
  JOURNAL   Plant Mol. Biol. 54 (3), 405-414 (2004)
   PUBMED   15284495
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .154
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="mRNA"
                     /isolate="Gifu B-129"
                     /db_xref="taxon:34305"
                     /clone="GENf085g08"
                     /tissue_type="mature nodule"
                     /clone_lib="Lotus corniculatus var. japonicus mature
                     nodule"
ORIGIN

Alignment Scores:
Pred. No.:                827             Length:         154
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       3               Gaps:           0

US-09-858-332G-28 (1-6) x BP061972 (1-154)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         29 CATAATCACAATCATAAT 46

RESULT 6
CE625685/c
LOCUS       CE625685                155 bp    DNA     linear   GSS 29-SEP-2003
DEFINITION  tigr-gss-dog-17000313653545 Dog Library Canis familiaris genomic,
            genomic survey sequence.
ACCESSION   CE625685
VERSION     CE625685.1  GI:36943060
KEYWORDS    GSS.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
            Canis.
REFERENCE   1  (bases 1 to 155)
  AUTHORS   Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
            Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
            Venter,J.C.
  TITLE     The dog genome: survey sequencing and comparative analysis
  JOURNAL   Science 301 (5641), 1898-1903 (2003)
   PUBMED   14512627
COMMENT     Contact: Kirkness EF
            The Institute for Genomic Research
            Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
            Rockville, MD 20850, USA
            Tel: 301-838-0200
            Fax: 301-838-0208
            Email: ekirknes@tigr.org
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .155
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /strain="Standard Poodle"
                     /db_xref="taxon:9615"
                     /clone_lib="Dog Library"
                     /note="Site_1: BstXI; Libraries were prepared from
                     peripheral blood"
ORIGIN

Alignment Scores:
Pred. No.:                832             Length:         155
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       10              Gaps:           0

US-09-858-332G-28 (1-6) x CE625685 (1-155)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        119 CATAATCATAATCATAAT 102

RESULT 7
AW584439/c
LOCUS       AW584439                156 bp    mRNA    linear   EST 07-SEP-2000
DEFINITION  N210487e MHAM Medicago truncatula/Glomus versiforme mixed EST
            library cDNA clone MHAM-2A14, mRNA sequence.
ACCESSION   AW584439
VERSION     AW584439.1  GI:7261493
KEYWORDS    EST.
SOURCE      Medicago truncatula/Glomus versiforme mixed EST library
  ORGANISM  Medicago truncatula/Glomus versiforme mixed EST library
            Eukaryota; mixed EST libraries.
REFERENCE   1  (bases 1 to 156)
  AUTHORS   Harrison,M.J., Liu,J., Peng,H., Gonzales,M., Ellis,L., Town,C.D.,
            Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.
  TITLE     ESTs from roots of Medicago truncatula after colonization with
            Glomus versiforme
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Harrison M.J.
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73401
            Tel: 580-223-5810
            Fax: 580-221-7380
            Email: mjharrison@noble.org
            Other name: MHAM-2b-A07; Date: 3/14/00; Updated to the Database of
            Expressed Sequence Tags (dbEST) on 04/27/00; More information is
            available at 'http://chrysie.tamu.edu/medicago'.
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .156
                     /organism="Medicago truncatula/Glomus versiforme mixed EST
                     library"
                     /mol_type="mRNA"
                     /cultivar="Medicago truncatula genotype A17"
                     /db_xref="taxon:119092"
                     /clone="MHAM-2A14"
                     /tissue_type="roots colonized with Glomus versiforme"
                     /dev_stage="Roots harvested at 10, 17, 22, 31 and 38 days
                     post-inoculation with Glomus versiforme. The library was
                     made from a mixture of RNA from each of these stages."
                     /lab_host="E. coli strain XLOLR"
                     /clone_lib="MHAM"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; cDNA was prepared from polyA+ enriched RNA from
                     roots harvested at 10, 17, 22, 31 and 38 days
                     post-inoculation with Glomus versiforme. The cDNA was
```

directionally ligated into the Unizap XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in XLOLR cells."
ORIGIN

Alignment Scores:

Pred. No.:	838	Length:	156
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0

US-09-858-332G-28 (1-6) x AW584439 (1-156)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         46 CACAATCACAATCACAAT 29
```

RESULT 8
CE340511
LOCUS CE340511 156 bp DNA linear GSS 26-SEP-2003
DEFINITION tigr-gss-dog-17000334081537 Dog Library Canis familiaris genomic, genomic survey sequence.
ACCESSION CE340511
VERSION CE340511.1 GI:36165047
KEYWORDS GSS.
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE 1 (bases 1 to 156)
AUTHORS Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K., Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and Venter,J.C.
TITLE The dog genome: survey sequencing and comparative analysis
JOURNAL Science 301 (5641), 1898-1903 (2003)
PUBMED 14512627
COMMENT Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirknes@tigr.org
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .156
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site_1: BstXI; Libraries were prepared from peripheral blood"
ORIGIN

Alignment Scores:

Pred. No.:	838	Length:	156
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	10	Gaps:	0

US-09-858-332G-28 (1-6) x CE340511 (1-156)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         75 CATAACCATAACCATAAC 92
```

RESULT 9
CO278529
LOCUS CO278529 157 bp mRNA linear EST 24-JUN-2004
DEFINITION EK130308.5prime Exelixis FlyTag CK01 pCDNA-SK+ Drosophila melanogaster cDNA clone EK130308 5, mRNA sequence.
ACCESSION CO278529
VERSION CO278529.1 GI:49199548
KEYWORDS EST.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 157)
AUTHORS Kopczynski,C., Platt,D., Campbell,J., Muzong,C., Laufer,A., Peterson,E. and Swimmer,C.
TITLE Exelixis FlyTag EST Project CK01 Library
JOURNAL Unpublished (2004)
COMMENT Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: EK.1303 row: A column: 8
High quality sequence stop: 155.
FEATURES Location/Qualifiers
source 1. .157
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="EK130308"
/clone_lib="Exelixis FlyTag CK01 pCDNA-SK+"
/note="Organ: mixed stage embryos, imaginal disks, and adult heads; Vector: pCDNA-SK+; Site_1: NotI; Site_2: XhoI; Random primed, normalized library from mixed stage embryos, imaginal disks, and adult heads."
ORIGIN

Alignment Scores:

Pred. No.:	844	Length:	157
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	7	Gaps:	0

US-09-858-332G-28 (1-6) x CO278529 (1-157)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         15 CACAACCACAACCACAAC 32
```

RESULT 10
CB214559
LOCUS CB214559 161 bp mRNA linear EST 05-FEB-2003
DEFINITION OML04839 Oryza minuta HybriZAP-2.1 XR library Oryza minuta cDNA 5', mRNA sequence.
ACCESSION CB214559
VERSION CB214559.1 GI:28260650
KEYWORDS EST.
SOURCE Oryza minuta
ORGANISM Oryza minuta
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 161)
AUTHORS Shin,J.S.
TITLE Oryza minuta HybriZAP-2.1 XR library
JOURNAL Unpublished (2003)
COMMENT Contact: Jeong Sheop Shin
Plant Molecular Genetics

```
           Graduate School of Biotechnology, University of Korea
           136-701 Anam-dong 5/1 Seoul, Korea
           Tel: 00 82 2 3290 3430
           Fax: 00 82 2 927 9028
           Email: jsshin@kuccnx.korea.ac.kr.
FEATURES             Location/Qualifiers
     source          1. .161
                     /organism="Oryza minuta"
                     /mol_type="mRNA"
                     /db_xref="taxon:63629"
                     /dev_stage="4-weeks after germination"
                     /clone_lib="Oryza minuta HybriZAP-2.1 XR library"
                     /note="Organ: immature leaf"
ORIGIN

Alignment Scores:
Pred. No.:              867             Length:         161
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x CB214559 (1-161)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         34 CACAACCACAACCACAAC 51

RESULT 11
BF645838
LOCUS       BF645838                163 bp    mRNA    linear   EST 20-DEC-2000
DEFINITION  NF042B06EC1F1048 Elicited cell culture Medicago truncatula cDNA
            clone NF042B06EC 5', mRNA sequence.
ACCESSION   BF645838
VERSION     BF645838.1  GI:11910967
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 163)
  AUTHORS   Torres-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,
            Flores,H.R., Inman,J.T., Weller,J.W. and May,G.D.
  TITLE     Expressed Sequence Tags from the Samuel Roberts Noble Foundation -
            Center for Medicago Genomics Research
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Dixon RA
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA
            Tel: 580 221 7302
            Fax: 580 221 7380
            Email: radixon@noble.org
            Insert Length: 163   Std Error: 0.00
            Plate: 042  row: B  column: 06
            Seq primer: TCACACAGGAAACAGCTATGAC.
FEATURES             Location/Qualifiers
     source          1. .163
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /db_xref="taxon:3880"
                     /clone="NF042B06EC"
                     /tissue_type="Cell cultures derived from root tissues"
                     /dev_stage="Cell suspensions were subcultured every 14
                     days. Cells were induced six days after subculture"
                     /clone_lib="Elicited cell culture"
                     /note="Vector: Lambda Zap; Cells were induced with yeast
                     cell wall extracts equivalent to  50ug/ml glucose in the
                     final concentration. Samples were taken at  0.5, 1, 12 and
                     24 hours after induction. Equal amounts of RNA from  each
                     time point were pooled and used for mRNA isolation."
ORIGIN

Alignment Scores:
Pred. No.:              878             Length:         163
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-858-332G-28 (1-6) x BF645838 (1-163)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        142 CATAATCATAATCATAAT 159

RESULT 12
BJ434601/c
LOCUS       BJ434601                163 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ434601 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv24j12 3', mRNA sequence.
ACCESSION   BJ434601
VERSION     BJ434601.1  GI:19409323
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 163)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .163
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv24j12"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:              878             Length:         163
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-858-332G-28 (1-6) x BJ434601 (1-163)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        149 CACAACCACAACCACAAC 132

RESULT 13
CV044762
LOCUS       CV044762                165 bp    mRNA    linear   EST 23-AUG-2004
DEFINITION  EST 6469 Green Apricot Fruit Lambda Zap II Library Prunus armeniaca
            cDNA clone bga009k03p 3', mRNA sequence.
ACCESSION   CV044762
VERSION     CV044762.1  GI:51502251
```

```
KEYWORDS    EST.
SOURCE      Prunus armeniaca (apricot)
  ORGANISM  Prunus armeniaca
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Rosales; Rosaceae; Amygdaloideae; Prunus.
REFERENCE   1  (bases 1 to 165)
  AUTHORS   Grimplet,J., Romieu,C., Audergon,J.M., Albagnac,G., Lambert,P.,
            Bouchet,J.P. and Terrier,N.
  TITLE     High Throughput Detection of Isogenes among 5724 3' EST from
            Apricot Fruit(Prunus armeniaca)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Audergon JM
            Unite de genetique et amelioration des fruits et legumes
            Institut National de la Recherche Agronomique
            Domaine Saint-Maurice BP 94 84143 Montfavet cedex
            Tel: 00-33-(0)4-32-72-26-68
            Fax: 00-33-(0)4-32-72-26-62
            Email: audergon@avignon.inra.fr
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .165
                     /organism="Prunus armeniaca"
                     /mol_type="mRNA"
                     /cultivar="Bergeron"
                     /db_xref="taxon:36596"
                     /clone="bga009k03p"
                     /dev_stage="Green stage"
                     /clone_lib="Green Apricot Fruit Lambda Zap II Library"
                     /note="Organ: Fruit; Vector: Lambda Zap II; Site_1: Eco
                     RI; Site_2: XhoI; Oriented library, construction described
                     in Molecular cloning and expression of a cDNA encoding
                     1-aminocyclopropane-1-carboxylate (ACC) oxidase from
                     apricot fruit (Prunus armeniaca cv. Bergeron) by
                     Mbeguie-Mbeguie D, Chahine H, Gomez RM, Gouble B, Audergon
                     JM, Souty M, Albagnac G, Fils-Lycaon B in Physiol Plant
                     105:294-303 1999"
ORIGIN

Alignment Scores:
Pred. No.:                890           Length:         165
Score:                    42.00         Matches:        6
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              100.00%       Indels:         0
DB:                       7             Gaps:           0

US-09-858-332G-28 (1-6) x CV044762 (1-165)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         96 CACAATCACAATCATAAC 113

RESULT 14
CV046698/c
LOCUS       CV046698                165 bp    mRNA    linear   EST 23-AUG-2004
DEFINITION  EST 8405 Green Apricot Fruit Lambda Zap II Library Prunus armeniaca
            cDNA clone bga009k03q 5', mRNA sequence.
ACCESSION   CV046698
VERSION     CV046698.1  GI:51504187
KEYWORDS    EST.
SOURCE      Prunus armeniaca (apricot)
  ORGANISM  Prunus armeniaca
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Rosales; Rosaceae; Amygdaloideae; Prunus.
REFERENCE   1  (bases 1 to 165)
  AUTHORS   Grimplet,J., Romieu,C., Audergon,J.M., Albagnac,G., Lambert,P.,
            Bouchet,J.P. and Terrier,N.
  TITLE     High Throughput Detection of Isogenes among 5724 3' EST from
            Apricot Fruit(Prunus armeniaca)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Audergon JM
            Unite de genetique et amelioration des fruits et legumes
            Institut National de la Recherche Agronomique
            Domaine Saint-Maurice BP 94 84143 Montfavet cedex
            Tel: 00-33-(0)4-32-72-26-68
            Fax: 00-33-(0)4-32-72-26-62
            Email: audergon@avignon.inra.fr
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .165
                     /organism="Prunus armeniaca"
                     /mol_type="mRNA"
                     /cultivar="Bergeron"
                     /db_xref="taxon:36596"
                     /clone="bga009k03q"
                     /dev_stage="Green stage"
                     /clone_lib="Green Apricot Fruit Lambda Zap II Library"
                     /note="Organ: Fruit; Vector: Lambda Zap II; Site_1: Eco
                     RI; Site_2: XhoI; Oriented library, construction described
                     in Molecular cloning and expression of a cDNA encoding
                     1-aminocyclopropane-1-carboxylate (ACC) oxidase from
                     apricot fruit (Prunus armeniaca cv. Bergeron) by
                     Mbeguie-Mbeguie D, Chahine H, Gomez RM, Gouble B, Audergon
                     JM, Souty M, Albagnac G, Fils-Lycaon B in Physiol Plant
                     105:294-303 1999"
ORIGIN

Alignment Scores:
Pred. No.:                890           Length:         165
Score:                    42.00         Matches:        6
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              100.00%       Indels:         0
DB:                       7             Gaps:           0

US-09-858-332G-28 (1-6) x CV046698 (1-165)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         70 CACAATCACAATCATAAC 53

RESULT 15
BJ429957/c
LOCUS       BJ429957                174 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ429957 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv5d17 3', mRNA sequence.
ACCESSION   BJ429957
VERSION     BJ429957.1  GI:19404679
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 174)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .174
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv5d17"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
```

```
ORIGIN

Alignment Scores:
Pred. No.:              941           Length:        174
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     3             Gaps:          0

US-09-858-332G-28 (1-6) x BJ429957 (1-174)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        127 CACAATCACAATCACAAT 110

RESULT 16
CO269710
LOCUS       CO269710                174 bp    mRNA    linear   EST 24-JUN-2004
DEFINITION  EK095625.5prime Exelixis FlyTag CK01 pCDNA-SK+ Drosophila
            melanogaster cDNA clone EK095625 5, mRNA sequence.
ACCESSION   CO269710
VERSION     CO269710.1  GI:49190728
KEYWORDS    EST.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 174)
  AUTHORS   Kopczynski,C., Platt,D., Campbell,J., Muzong,C., Laufer,A.,
            Peterson,E. and Swimmer,C.
  TITLE     Exelixis FlyTag EST Project CK01 Library
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Based upon one or more reads of this clone where vector sequence
            was present at both ends, this clone has been determined to contain
            contain a cDNA insert on the order of 600-1000 bases.
            Plate: EK.956  row: C  column: 1
            High quality sequence stop: 173.
FEATURES             Location/Qualifiers
     source          1. .174
                     /organism="Drosophila melanogaster"
                     /mol_type="mRNA"
                     /db_xref="taxon:7227"
                     /clone="EK095625"
                     /clone_lib="Exelixis FlyTag CK01 pCDNA-SK+"
                     /note="Organ: mixed stage embryos, imaginal disks, and
                     adult heads; Vector: pCDNA-SK+; Site_1: NotI; Site_2:
                     XhoI; Random primed, normalized library from mixed stage
                     embryos, imaginal disks, and adult heads."
ORIGIN

Alignment Scores:
Pred. No.:              941           Length:        174
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     7             Gaps:          0

US-09-858-332G-28 (1-6) x CO269710 (1-174)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        140 CACAACCATAATCACAAT 157

RESULT 17
CV583709
LOCUS       CV583709                184 bp    mRNA    linear   EST 22-OCT-2004
DEFINITION  L_AH-aaa11c08.b1 Hc7B_M Ajellomyces capsulatus cDNA 3', mRNA
            sequence.
ACCESSION   CV583709
VERSION     CV583709.1  GI:54571298
KEYWORDS    EST.
SOURCE      Ajellomyces capsulatus
  ORGANISM  Ajellomyces capsulatus
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Onygenales; Ajellomycetaceae; Ajellomyces.
REFERENCE   1  (bases 1 to 184)
  AUTHORS   Magrini,V., Sil,A., Goldman,W. and Mardis,R.
  TITLE     Genomic Resources for Histoplasma Sequencing Mardis/WashU
            Histoplasma capsulatum Sequencing Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Vincent Magrini
            Mardis/Goldman/Sil Genomic Resources for Histoplasma Sequencing
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: vmagrini@watson.wustl.edu
            Contact Vincent Magrini (vmagrini@watson.wustl.edu) for further
            information relating to organism, libraries, or clone availability.
            Seq primer: M13_reverse.
FEATURES             Location/Qualifiers
     source          1. .184
                     /organism="Ajellomyces capsulatus"
                     /mol_type="mRNA"
                     /strain="G217B"
                     /db_xref="taxon:5037"
                     /dev_stage="Mycelia"
                     /lab_host="GC10"
                     /clone_lib="Hc7B_M"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI flanks TA cloning
                     site; Site_2: EcoRI flanks TA cloning site; polyA mRNA was
                     captured with oligo d(T) magnetic beads and eluted.
                     AAGCAGTGGTAACAACGCAGAGTACTTTTTTTTTTTTTTTVN was used for
                     priming reverse transcription in the presence of
                     SuperScript II RTase. The 5' primer (5'
                     GCGGCCGCCTAATACGACTCACTATAGGG) annealed to cytosine
                     nucleosides added by terminal transferase to the 5' end of
                     the transcript. Amplification during the PCR used primers
                     5' AAGCAGTGGTAACAACGCAGA and 5' GGCCGCTAATACGACTCACTATAGGG
                     and Clone-TECH Taq DNA polymerase. PCR products between
                     500 and 2,000 bp were size-fractioned from 0.7% LE/TAE
                     agarose and cloned into the Topo TA site of Invitrogen
                     plasmid vector pCR4-TOPO."
ORIGIN

Alignment Scores:
Pred. No.:              999           Length:        184
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     7             Gaps:          0

US-09-858-332G-28 (1-6) x CV583709 (1-184)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         92 CATAATCATAATCATAAT 109

RESULT 18
CV583842/c
LOCUS       CV583842                185 bp    mRNA    linear   EST 22-OCT-2004
DEFINITION  L_AH-aaa11c08.g1 Hc7B_M Ajellomyces capsulatus cDNA 5', mRNA
            sequence.
ACCESSION   CV583842
VERSION     CV583842.1  GI:54571431
KEYWORDS    EST.
```

```
SOURCE      Ajellomyces capsulatus
  ORGANISM  Ajellomyces capsulatus
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Onygenales; Ajellomycetaceae; Ajellomyces.
REFERENCE   1  (bases 1 to 185)
  AUTHORS   Magrini,V., Sil,A., Goldman,W. and Mardis,R.
  TITLE     Genomic Resources for Histoplasma Sequencing Mardis/WashU
            Histoplasma capsulatum Sequencing Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Vincent Magrini
            Mardis/Goldman/Sil Genomic Resources for Histoplasma Sequencing
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: vmagrini@watson.wustl.edu
            Contact Vincent Magrini (vmagrini@watson.wustl.edu) for further
            information relating to organism, libraries, or clone availability.
            Seq primer: M13_forward.
FEATURES             Location/Qualifiers
     source          1. .185
                     /organism="Ajellomyces capsulatus"
                     /mol_type="mRNA"
                     /strain="G217B"
                     /db_xref="taxon:5037"
                     /dev_stage="Mycelia"
                     /lab_host="GC10"
                     /clone_lib="Hc7B_M"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI flanks TA cloning
                     site; Site_2: EcoRI flanks TA cloning site; polyA mRNA was
                     captured with oligo d(T) magnetic beads and eluted.
                     AAGCAGTGGTAACAACGCAGAGTACTTTTTTTTTTTTTTTVN was used for
                     priming reverse transcription in the presence of
                     SuperScript II RTase. The 5' primer (5'
                     GCGGCCGCCTAATACGACTCACTATAGGG) annealed to cytosine
                     nucleosides added by terminal transferase to the 5' end of
                     the transcript. Amplification during the PCR used primers
                     5' AAGCAGTGGTAACAACGCAGA and 5' GGCCGCTAATACGACTCACTATAGGG
                     and Clone-TECH Taq DNA polymerase. PCR products between
                     500 and 2,000 bp were size-fractioned from 0.7% LE/TAE
                     agarose and cloned into the Topo TA site of Invitrogen
                     plasmid vector pCR4-TOPO."
ORIGIN

Alignment Scores:
Pred. No.:                1e+03          Length:         185
Score:                    42.00          Matches:        6
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       7              Gaps:           0

US-09-858-332G-28 (1-6) x CV583842 (1-185)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         93 CATAATCATAATCATAAT 76

RESULT 19
BF202384
LOCUS       BF202384                193 bp    mRNA    linear   EST 06-NOV-2000
DEFINITION  WHE1764_A06_A12ZS Wheat pre-anthesis spike cDNA library Triticum
            aestivum cDNA clone WHE1764_A06_A12, mRNA sequence.
ACCESSION   BF202384
VERSION     BF202384.1  GI:11117126
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 193)
  AUTHORS   Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D.,
            Han,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J.,
            Seaton,C.L. and Tong,J.C.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Pre-anthesis spike cDNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene SK primer.
FEATURES             Location/Qualifiers
     source          1. .193
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /cultivar="Chinese Spring"
                     /db_xref="taxon:4565"
                     /clone="WHE1764_A06_A12"
                     /tissue_type="Spike before anthesis"
                     /dev_stage="Adult plant"
                     /lab_host="E. coli SOLR"
                     /clone_lib="Wheat pre-anthesis spike cDNA library"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; Plants were grown in the
                     greenhouse.  Whole spike with awns trimmed,  white, green
                     and yellow anther were collected and total RNA, and
                     poly(A) RNA were prepared, a cDNA library was made, and
                     the cDNA  clones were in vivo excised to give pBluescript
                     phagemids in the TJ  Close lab (Choi, Close, Fenton) at
                     the University of California,  Riverside.  Plasmid DNA
                     preparations and DNA sequencing were performed  in the OD
                     Anderson lab (all other authors)."
ORIGIN

Alignment Scores:
Pred. No.:                1.05e+03       Length:         193
Score:                    42.00          Matches:        6
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       2              Gaps:           0

US-09-858-332G-28 (1-6) x BF202384 (1-193)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        157 CACAACCACAACCACAAT 174

RESULT 20
AI577627
LOCUS       AI577627                196 bp    mRNA    linear   EST 05-APR-1999
DEFINITION  UI-R-AB0-vr-c-05-0-UI.s1 UI-R-AB0 Rattus norvegicus cDNA clone
            UI-R-AB0-vr-c-05-0-UI 3', mRNA sequence.
ACCESSION   AI577627
VERSION     AI577627.1  GI:4562003
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 196)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  PUBMED    8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
```

```
            University of Iowa
            375 Newton Road , 4156  MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            clones will be available through Research Genetics (www.resgen.com)
            Seq primer: M13 Forward.
FEATURES              Location/Qualifiers
     source           1. .196
                      /organism="Rattus norvegicus"
                      /mol_type="mRNA"
                      /strain="Sprague-Dawley"
                      /db_xref="taxon:10116"
                      /clone="UI-R-AB0-vr-c-05-0-UI"
                      /dev_stage="adult"
                      /lab_host="DH10B (Life Technologies)"
                      /clone_lib="UI-R-AB0"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-AB0
                      library is a non-normalized library constructed from  16.5
                      dpc rat ventricle.  The tag is a string of 5 nucleotides
                      present  between the Not I site and the oligo-dT track.
                      The library was  constructed as described by Bonaldo,
                      Lennon and Soares,  Genome Research 6: 791-806, 1996.
                      Tissue provided by Jim Lin, Department of Biology,
                      University of Iowa."
ORIGIN

Alignment Scores:
Pred. No.:                1.07e+03        Length:          196
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       1               Gaps:            0

US-09-858-332G-28 (1-6) x AI577627 (1-196)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         40 CATAATCACAACCATAAC 57

RESULT 21
BQ855785/c
LOCUS       BQ855785                 204 bp    mRNA    linear   EST 14-AUG-2002
DEFINITION  QGB27L15.yg.ab1 QG_ABCDI lettuce salinas Lactuca sativa cDNA clone
            QGB27L15, mRNA sequence.
ACCESSION   BQ855785
VERSION     BQ855785.1  GI:22241250
KEYWORDS    EST.
SOURCE      Lactuca sativa
  ORGANISM  Lactuca sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
            Cichorieae; Lactuca.
REFERENCE   1  (bases 1 to 204)
  AUTHORS   Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J.,
            Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y.,
            Lai,Z., Church,S., Jackson,L. and Bradford,K.
  TITLE     Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
            singleton, see http://cgpdb.ucdavis.edu/ for details.
            Plate: QGB27  row: L  column: 15.
FEATURES              Location/Qualifiers
     source           1. .204
                      /organism="Lactuca sativa"
                      /mol_type="mRNA"
                      /cultivar="Salinas"
                      /db_xref="taxon:4236"
                      /clone="QGB27L15"
                      /lab_host="E.coli"
                      /clone_lib="QG_ABCDI lettuce salinas"
                      /note="Vector: pBRcDNASfiAB; The library was constructed
                      from 10 different sources of RNA from a single genotype.
                      Separate cDNAs were generated using primers that
                      incorporated unique 5' and 3' tags to distinguish each
                      source of RNA. cDNAs were then pooled, size-fractionated,
                      directionally cloned into a custom medium-copy vector and
                      transformations made with four size classes to minimize
                      size bias. Details of each source of RNA and library
                      construction can be obtained at http://cgpdb.ucdavis.edu/
                      TAG_TISSUE=germinating seeds
                      TAG_LIB=QG_ABCDI lettuce salinas
                      TAG_SEQ=TCTGTGCGGG"
ORIGIN

Alignment Scores:
Pred. No.:                1.11e+03        Length:          204
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       5               Gaps:            0

US-09-858-332G-28 (1-6) x BQ855785 (1-204)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        144 CATAATCATAATCATAAT 127

RESULT 22
CB637138/c
LOCUS       CB637138                 208 bp    mRNA    linear   EST 08-APR-2003
DEFINITION  OSJNEa03F22.f OSJNEa Oryza sativa (japonica cultivar-group) cDNA
            clone OSJNEa03F22 5', mRNA sequence.
ACCESSION   CB637138
VERSION     CB637138.1  GI:29632129
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 208)
  AUTHORS   Jantasuriyarat,C., Gowda,M., Haller,K., Hatfield,J., Lu,G.,
            Stahlberg,E., Zhou,B., Li,H., Kim,H., Yu,Y., Dean,R.A., Wing,R.A.,
            Soderlund,C. and Wang,G.L.
  TITLE     Large-scale identification of expressed sequence tags involved in
            rice and rice blast fungus interaction
  JOURNAL   Plant Physiol. 138 (1), 105-115 (2005)
   PUBMED   15888683
COMMENT     Contact: Rod Wing
            Arizona Genomics Institute
            University of Arizona
            Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
            85721-0088, USA
            Tel: 520 626 3967
            Fax: 520 621 9288
            Email: rwing@genome.arizona.edu
            PCR PRimers
            FORWARD: atc agc ggc cgc gat cc
            BACKWARD: aat taa ccc tca cta aag gg
```

```
           Plate: 03  row: F  column: 22
           Seq primer: atc agc ggc cgc gat cc.
FEATURES             Location/Qualifiers
     source          1. .208
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="OSJNEa03F22"
                     /tissue_type="Leaf"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
                     /clone_lib="OSJNEa"
                     /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
                     XhoI; 6 hrs after innoculation with Rice Blast (Che
                     86061)"
ORIGIN

Alignment Scores:
Pred. No.:              1.14e+03       Length:        208
Score:                  42.00          Matches:       6
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     6              Gaps:          0

US-09-858-332G-28 (1-6) x CB637138 (1-208)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        206 CACAACCACAACCACAAC 189

RESULT 23
AU053723
LOCUS       AU053723                209 bp    mRNA    linear   EST 28-APR-1999
DEFINITION  AU053723 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
            discoideum cDNA clone SLJ555, mRNA sequence.
ACCESSION   AU053723
VERSION     AU053723.1  GI:4702205
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 209)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp
            PROJECT = Dictyostelium discoideum cDNA project in Japan.
FEATURES             Location/Qualifiers
     source          1. .209
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SLJ555"
                     /dev_stage="slug"
                     /clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
ORIGIN

Alignment Scores:
Pred. No.:              1.14e+03       Length:        209
Score:                  42.00          Matches:       6
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     1              Gaps:          0

US-09-858-332G-28 (1-6) x AU053723 (1-209)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db          2 CACAACCACAACCACAAC 19

RESULT 24
BQ852441/c
LOCUS       BQ852441                209 bp    mRNA    linear   EST 14-AUG-2002
DEFINITION  QGB18A16.yg.ab1 QG_ABCDI lettuce salinas Lactuca sativa cDNA clone
            QGB18A16, mRNA sequence.
ACCESSION   BQ852441
VERSION     BQ852441.1  GI:22237906
KEYWORDS    EST.
SOURCE      Lactuca sativa
  ORGANISM  Lactuca sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
            Cichorieae; Lactuca.
REFERENCE   1  (bases 1 to 209)
  AUTHORS   Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J.,
            Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y.,
            Lai,Z., Church,S., Jackson,L. and Bradford,K.
  TITLE     Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
            singleton, see http://cgpdb.ucdavis.edu/ for details.
            Plate: QGB18  row: A  column: 16.
FEATURES             Location/Qualifiers
     source          1. .209
                     /organism="Lactuca sativa"
                     /mol_type="mRNA"
                     /cultivar="Salinas"
                     /db_xref="taxon:4236"
                     /clone="QGB18A16"
                     /lab_host="E.coli"
                     /clone_lib="QG_ABCDI lettuce salinas"
                     /note="Vector: pBRcDNASfiAB; The library was constructed
                     from 10 different sources of RNA from a single genotype.
                     Separate cDNAs were generated using primers that
                     incorporated unique 5' and 3' tags to distinguish each
                     source of RNA. cDNAs were then pooled, size-fractionated,
                     directionally cloned into a custom medium-copy vector and
                     transformations made with four size classes to minimize
                     size bias. Details of each source of RNA and library
                     construction can be obtained at http://cgpdb.ucdavis.edu/
                     TAG_SEQ=Not found"
ORIGIN

Alignment Scores:
Pred. No.:              1.14e+03       Length:        209
Score:                  42.00          Matches:       6
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     5              Gaps:          0

US-09-858-332G-28 (1-6) x BQ852441 (1-209)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
```

```
Db        207 CATAATCATAATCATAAT 190

RESULT 25
AK185713
LOCUS       AK185713                211 bp    mRNA    linear   HTC 23-NOV-2004
DEFINITION  Mus musculus cDNA, clone:Y0G0134P19, strand:plus,
            reference:ENSEMBL:Mouse-Transcript-ENST:ENSMUST00000053540, based
            on BLAT search.
ACCESSION   AK185713
VERSION     AK185713.1  GI:56009890
KEYWORDS    HTC; ASSETS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Watahiki,A., Waki,K., Hayatsu,N., Shiraki,T., Kondo,S.,
            Nakamura,M., Sasaki,D., Arakawa,T., Kawai,J., Harbers,M.,
            Hayashizaki,Y. and Carninci,P.
  TITLE     Libraries enriched for alternatively spliced exons reveal splicing
            patterns in melanocytes and melanomas
  JOURNAL   Nat. Methods 1, 233-239 (2004)
REFERENCE   2  (bases 1 to 211)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Harbers,M., Hayatsu,N.,
            Hori,F., Imotani,K., Kawai,J., Kondo,S., Murata,M., Nakamura,M.,
            Nomura,K., Ohno,M., Sasaki,D., Shiraki,T., Waki,K., Watahiki,A. and
            Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-2004) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa, 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
            URL:http://genome.gsc.riken.jp, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Alternative Splicing Libraries (ASLs) are prepared by: Preparating
            of single-stranded DNA using a RNA template from full length cDNA
            libraries, hybridizating of
            single-stranded DNAs, removing of remaining single-stranded DNA,
            digesting of regions comprising double-stranded DNA by a set of 4
            bp-cutters, capturing of DNA hybrids with loop structures
            (alternative spliced exon), ligating of Y-shaped primers to
            isolated DNA hybrids with loop structures, PCR amplification of
            ligation products and their cloning into pFLCI vector. (Reference).
FEATURES             Location/Qualifiers
     source          1. .211
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="Y0G0134P19"
                     /cell_line="mixture of B16-F10Y and melan-c"
                     /cell_type="mixture of melanoma cell and melanocyte cell"
                     /clone_lib="Alternative Splicing Library L2"
                     /note="strand:plus, reference:ENSEMBL:Mouse-Transcript-
                     ENST:ENSMUST00000053540, based on BLAT search"
ORIGIN

Alignment Scores:
Pred. No.:              1.16e+03        Length:         211
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     4               Gaps:           0

US-09-858-332G-28 (1-6) x AK185713 (1-211)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         58 CACAACCATAACCACAAC 75

RESULT 26
BQ586973
LOCUS       BQ586973                214 bp    mRNA    linear   EST 06-DEC-2002
DEFINITION  E012352-024-011-P08-SP6 MPIZ-ADIS-024-leaf Beta vulgaris cDNA clone
            024-011-P08 5-PRIME, mRNA sequence.
ACCESSION   BQ586973
VERSION     BQ586973.1  GI:26116555
KEYWORDS    EST.
SOURCE      Beta vulgaris
  ORGANISM  Beta vulgaris
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            Caryophyllales; Amaranthaceae; Beta.
REFERENCE   1  (bases 1 to 214)
  AUTHORS   Herwig,R., Schulz,B., Weisshaar,B., Hennig,S., Steinfath,M.,
            Drungowski,M., Stahl,D., Wruck,W., Menze,A., O'Brien,J., Lehrach,H.
            and Radelof,U.
  TITLE     Construction of a 'unigene' cDNA clone set by oligonucleotide
            fingerprinting allows access to 25 000 potential sugar beet genes
  JOURNAL   Plant J. 32 (5), 845-857 (2002)
  PUBMED    12472698
COMMENT     Contact: Weisshaar B
            ADIS DNA core facility at MPIZ
            Max-Planck-Institute for Plant Breeding Research
            Carl-von-Linne Weg 10, 50829 Koeln, Germany
            Fax: 00492215062851
            Email: weisshaa@mpiz-koeln.mpg.de
            Insert Length: 214   Std Error: 0.00
            Plate: 11  row: P  column: 08
            Seq primer: SP6; CATACGATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .214
                     /organism="Beta vulgaris"
                     /mol_type="mRNA"
                     /cultivar="KWS2320 (double haploid, monogerm breeding
                     line)"
                     /db_xref="GABI:185888"
                     /db_xref="taxon:161934"
                     /clone="024-011-P08"
                     /tissue_type="leaf"
                     /lab_host="EMDH10B"
                     /clone_lib="MPIZ-ADIS-024-leaf"
                     /note="Vector: pCMVSPORT6; Site_1: SalI; Site_2: NotI;
                     cDNA library from sugar beet, library provided by KWS
                     Kleinwanzlebener Saatzucht AG Einbeck, Germany, contact:
                     b.schulz@kws.de; cloning sites SalI-NotI, primer sites and
                     orientation:
                     SP6-SalI-CCACGCGTCCG-5prime-cDNA-polyA-CC-NotI-T7;  Note:
                     Sequencing granted in the context of the GABI-Beet
                     project, local PI: Dr. Katharina Schneider, coordinator:
                     Prof. Christian Jung;  Sequence submission managed by
                     RZPD/GABI-Primary database:http://gabi.rzpd.de"
ORIGIN

Alignment Scores:
Pred. No.:              1.17e+03        Length:         214
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     5               Gaps:           0

US-09-858-332G-28 (1-6) x BQ586973 (1-214)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        169 CACAACCACAACCACAAC 186

RESULT 27
AW158394
LOCUS       AW158394                216 bp    mRNA    linear   EST 05-NOV-1999
DEFINITION  za40h06.x1 Xenopus EST library Xenopus laevis cDNA clone za40h06
            5', mRNA sequence.
ACCESSION   AW158394
```

```
VERSION     AW158394.1  GI:6270423
KEYWORDS    EST.
SOURCE      Xenopus laevis (African clawed frog)
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Xenopus.
REFERENCE   1  (bases 1 to 216)
  AUTHORS   Schutz,K., de la Bastide,M., Huang,E.N., Nascimento,L., Preston,R.,
            Shah,R., Swaby,I., Shekher,M., Spiegel,L., Vil,M.D. and
            McCombie,W.R.
  TITLE     Expressed sequence tags from Xenopus
  JOURNAL   Unpublished (1999)
COMMENT     Contact: W. Richard McCombie
            Lita Annenberg Hazen Genome Sequencing Center
            Cold Spring Harbor Laboratory
            PO Box 100, Cold Spring Harbor, NY 11724, USA
            Tel: 516 367 8884
            Fax: 516 367 8874
            Email: mccombie@cshl.org
            Plate: za40  row: h  column: 06
            Seq primer: M13 universal forward primer
            High quality sequence stop: 216.
FEATURES             Location/Qualifiers
     source          1. .216
                     /organism="Xenopus laevis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8355"
                     /clone="za40h06"
                     /tissue_type="total brain tissue"
                     /cell_line="W22-TGA"
                     /dev_stage="tadpole"
                     /clone_lib="Xenopus EST library"
                     /note="Vector: Lambda Zap I; Site_1: XbaI; This library
                     was supplied by Holly Cline (Cold Spring Harbor Labs).
                     cDNA synthesis with oligo dT Xba I (Xba I cloning site).
                     RNA: stage 50-56 tadpoles, total brain tissue, GTC
                     extraction method."
ORIGIN

Alignment Scores:
Pred. No.:              1.18e+03      Length:         216
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     1             Gaps:           0

US-09-858-332G-28 (1-6) x AW158394 (1-216)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        105 CATAATCACAACCACAAT 122

RESULT 28
CB214364
LOCUS       CB214364                220 bp    mRNA    linear   EST 05-FEB-2003
DEFINITION  OML04644 Oryza minuta HybriZAP-2.1 XR library Oryza minuta cDNA 5',
            mRNA sequence.
ACCESSION   CB214364
VERSION     CB214364.1  GI:28260455
KEYWORDS    EST.
SOURCE      Oryza minuta
  ORGANISM  Oryza minuta
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 220)
  AUTHORS   Shin,J.S.
  TITLE     Oryza minuta HybriZAP-2.1 XR library
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Jeong Sheop Shin
            Plant Molecular Genetics
            Graduate School of Biotechnology, University of Korea
            136-701 Anam-dong 5/1 Seoul, Korea
            Tel: 00 82 2 3290 3430
            Fax: 00 82 2 927 9028
            Email: jsshin@kuccnx.korea.ac.kr.
FEATURES             Location/Qualifiers
     source          1. .220
                     /organism="Oryza minuta"
                     /mol_type="mRNA"
                     /db_xref="taxon:63629"
                     /dev_stage="4-weeks after germination"
                     /clone_lib="Oryza minuta HybriZAP-2.1 XR library"
                     /note="Organ: immature leaf"
ORIGIN

Alignment Scores:
Pred. No.:              1.21e+03      Length:         220
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     6             Gaps:           0

US-09-858-332G-28 (1-6) x CB214364 (1-220)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         93 CACAACCACAACCACAAC 110

RESULT 29
R44827
LOCUS       R44827                  222 bp    mRNA    linear   EST 22-MAY-1995
DEFINITION  yg31c02.s1 Soares infant brain 1NIB Homo sapiens cDNA clone
            IMAGE:34053 3', mRNA sequence.
ACCESSION   R44827
VERSION     R44827.1  GI:824203
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 222)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     On May 9, 1995 this sequence version replaced gi:803551.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1602
            High quality sequence stops: 132 Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1602   Std Error: 0.00
            Seq primer: Promega -21m13
            High quality sequence stop: 132.
FEATURES             Location/Qualifiers
     source          1. .222
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="GDB:406400"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:34053"
                     /sex="female"
                     /dev_stage="73 days post natal"
```

/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares infant brain 1NIB"
/note="Organ: whole brain; Vector: Lafmid BA; Site_1: Not I; Site_2: Hind III; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Hind III adaptors (Pharmacia), digested with Not I and directionally cloned into the Not I and Hind III sites of the Lafmid BA vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."
ORIGIN

Alignment Scores:

Pred. No.:	1.22e+03	Length:	222
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	8	Gaps:	0

US-09-858-332G-28 (1-6) x R44827 (1-222)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        202 CATAATCATAATCATAAT 219
```

RESULT 30
CO282440/c
LOCUS CO282440 223 bp mRNA linear EST 24-JUN-2004
DEFINITION EK160504.5prime Exelixis FlyTag CK01 pCDNA-SK+ Drosophila melanogaster cDNA clone EK160504 5, mRNA sequence.
ACCESSION CO282440
VERSION CO282440.1 GI:49203767
KEYWORDS EST.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 223)
AUTHORS Kopczynski,C., Platt,D., Campbell,J., Muzong,C., Laufer,A., Peterson,E. and Swimmer,C.
TITLE Exelixis FlyTag EST Project CK01 Library
JOURNAL Unpublished (2004)
COMMENT Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Based upon one or more reads of this clone where vector sequence was present at both ends, this clone has been determined to contain contain a cDNA insert on the order of 600-1000 bases.
Plate: EK.1605 row: A column: 4
High quality sequence stop: 220.
FEATURES Location/Qualifiers
source 1. .223
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="EK160504"
/clone_lib="Exelixis FlyTag CK01 pCDNA-SK+"
/note="Organ: mixed stage embryos, imaginal disks, and adult heads; Vector: pCDNA-SK+; Site_1: NotI; Site_2: XhoI; Random primed, normalized library from mixed stage embryos, imaginal disks, and adult heads."
ORIGIN

Alignment Scores:

Pred. No.:	1.23e+03	Length:	223
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	7	Gaps:	0

US-09-858-332G-28 (1-6) x CO282440 (1-223)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        137 CATAATCATAATCATAAT 120
```

RESULT 31
CE383475/c
LOCUS CE383475 224 bp DNA linear GSS 27-SEP-2003
DEFINITION tigr-gss-dog-17000334236131 Dog Library Canis familiaris genomic, genomic survey sequence.
ACCESSION CE383475
VERSION CE383475.1 GI:36616359
KEYWORDS GSS.
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE 1 (bases 1 to 224)
AUTHORS Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K., Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and Venter,J.C.
TITLE The dog genome: survey sequencing and comparative analysis
JOURNAL Science 301 (5641), 1898-1903 (2003)
PUBMED 14512627
COMMENT Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirknes@tigr.org
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .224
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site_1: BstXI; Libraries were prepared from peripheral blood"
ORIGIN

Alignment Scores:

Pred. No.:	1.23e+03	Length:	224
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	10	Gaps:	0

US-09-858-332G-28 (1-6) x CE383475 (1-224)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         28 CACAATCATAATCATAAC 11
```

RESULT 32
CO484355
LOCUS CO484355 228 bp mRNA linear EST 29-JUN-2005
DEFINITION GQ0204.B3_E17 GQ020: Clean ROOTS systems -Diurnal cycle Picea glauca cDNA clone GenomeQuebec_Id:GQ0204E17 5', mRNA sequence.
ACCESSION CO484355
VERSION CO484355.1 GI:50167736
KEYWORDS EST.
SOURCE Picea glauca (white spruce)
ORGANISM Picea glauca

```
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
REFERENCE   1  (bases 1 to 228)
  AUTHORS   Morency,M.-J., Cooke,J., Pavy,N., Parsons,L., Paule,C., Seguin,A.,
            Retzel,E., Butterfield,Y., Barber,S., Yang,G., Stott,J.,
            Siddiqui,A., Holt,R., Marra,M. and MacKay,J.
  TITLE     Arborea EST sequencing in Picea glauca (white spruce)
  JOURNAL   Unpublished (2004)
COMMENT     Contact: John MacKay
            Centre de Recherche en Biologie Forestiere
            Universite Laval
            Pavillon Charles-Eugene Marchand, Quebec, Quebec, CANADA G1K 7P4
            Fax: 418 656 7493
            Email: jmackay@rsvs.ulaval.ca
            Center for Computational Genomics and Bioinformatics (CCGB),
            University of Minnesota, MN_id Identifier: MN5235471 Clone ID:
            GQ0204_E17 Clones available through: John MacKay, Ph. D. Professeur
            adjoint -Assistant professor EMAIL: jmackay@rsvs.ulaval.ca Centre
            de Recherche en Biologie Forestiere (Forest Biology Research
            Center) Universite Laval Quebec, Quebec CANADA G1K 7P4
            Plate: 4  row: 17  column: E
            Seq primer: T3 Primer.
FEATURES             Location/Qualifiers
     source          1. .228
                     /organism="Picea glauca"
                     /mol_type="mRNA"
                     /strain="pg-653"
                     /db_xref="taxon:3330"
                     /clone="GenomeQuebec_Id:GQ0204E17"
                     /sex="Hermaphrodite"
                     /tissue_type="Whole roots systems, predominantly free of
                     michorizee"
                     /lab_host="E. coli DH10B cells"
                     /clone_lib="GQ020: Clean ROOTS systems -Diurnal cycle"
                     /note="Organ: whole root systems form 15 cm tall, 4 month
                     old seedlings produced by somatic embryogenesis; Vector:
                     pBluescript II SK (+) XR; Site_1: Eco-RI; Site_2: Xho-I;
                     Tissues were pooled from several plants grown under 20h
                     light/4h dark and collected every 2 hours over a period of
                     28 hours (4 dark time points and 11 time points in light).
                     cDNA was prepared from 5 mg of poly A+ selected RNA and
                     was directionally ligated into the pBluescript II SK (+)
                     XR vector (Stratagene), transformed by electroporation
                     into DH10B cells (In vitrogen) for propagation"
ORIGIN

Alignment Scores:
Pred. No.:              1.26e+03        Length:         228
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     7               Gaps:           0

US-09-858-332G-28 (1-6) x CO484355 (1-228)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         88 CACAACCACAACCACAAC 105

RESULT 33
AW689053
LOCUS       AW689053                231 bp    mRNA    linear   EST 15-JUN-2000
DEFINITION  NF014H03ST1F1000 Developing stem Medicago truncatula cDNA clone
            NF014H03ST 5', mRNA sequence.
ACCESSION   AW689053
VERSION     AW689053.1  GI:7563789
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 231)
  AUTHORS   He,X.-Z., Shadle,G., Scott,A.D., Harris,A.R., Gonzales,R.A.,
            Bell,C.J., Flores,H.R., Inman,J.T., Weller,J.W., May,G.D. and
            Dixon,R.A.
  TITLE     Expressed Sequence Tags from the Samuel Roberts Noble Foundation
            Medicago truncatula stem library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Dixon RA
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA
            Tel: 580 221 7302
            Fax: 580 221 7380
            Email: radixon@noble.org
            Insert Length: 231   Std Error: 0.00
            Plate: 014  row: H  column: 03
            Seq primer: TCACACAGGAAACAGCTATGAC.
FEATURES             Location/Qualifiers
     source          1. .231
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /db_xref="taxon:3880"
                     /clone="NF014H03ST"
                     /tissue_type="stem"
                     /dev_stage="Pooled developmental"
                     /clone_lib="Developing stem"
                     /note="Vector: Lambda Zap; Contains a mixture of
                     internodal stem segments"
ORIGIN

Alignment Scores:
Pred. No.:              1.27e+03        Length:         231
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AW689053 (1-231)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        100 CACAATCACAATCACAAC 117

RESULT 34
AU073601
LOCUS       AU073601                233 bp    mRNA    linear   EST 24-JUN-1999
DEFINITION  AU073601 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
            discoideum cDNA clone SSH644, mRNA sequence.
ACCESSION   AU073601
VERSION     AU073601.1  GI:5180022
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 233)
  AUTHORS   Urushihara,H.
  TITLE     Developmental cDNA in Dictyostelium discoideum (1999)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp
            PROJECT = Dictyostelium discoideum cDNA project in Japan.
FEATURES             Location/Qualifiers
     source          1. .233
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
```

```
                     /db_xref="taxon:44689"
                     /clone="SSH644"
                     /dev_stage="slug"
                     /clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
ORIGIN

Alignment Scores:
Pred. No.:                 1.28e+03        Length:          233
Score:                     42.00           Matches:         6
Percent Similarity:        100.00%         Conservative:    0
Best Local Similarity:     100.00%         Mismatches:      0
Query Match:               100.00%         Indels:          0
DB:                        1               Gaps:            0

US-09-858-332G-28 (1-6) x AU073601 (1-233)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        159 CACAACCACAACCACAAC 176

RESULT 35
AU302349/c
LOCUS       AU302349                 236 bp    mRNA    linear   EST 22-JAN-2004
DEFINITION  AU302349 zinnia cultured mesophyll cell equalized cDNA Zinnia
            elegans cDNA clone Z10877, mRNA sequence.
ACCESSION   AU302349
VERSION     AU302349.1  GI:41118208
KEYWORDS    EST.
SOURCE      Zinnia elegans
  ORGANISM  Zinnia elegans
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; campanulids; Asterales; Asteraceae; Asteroideae;
            Heliantheae; Zinnia.
REFERENCE   1  (bases 1 to 236)
  AUTHORS   Demura,T., Tashiro,G., Horiguchi,G., Kishimoto,N., Kubo,M.,
            Matsuoka,N., Minami,A., Nagata-Hiwatashi,M., Nakamura,K.,
            Okamura,Y., Sassa,N., Suzuki,S., Yazaki,J., Kikuchi,S. and
            Fukuda,H.
  TITLE     Visualization by comprehensive microarray analysis of gene
            expression programs during transdifferentiation of mesophyll cells
            into xylem cells
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (24), 15794-15799 (2002)
COMMENT     Contact: Taku Demura
            Morphogenesis Research Group
            RIKEN Plant Science Center
            1-7-22 Suehirocho, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9605
            Fax: 81-45-503-9573
            Email: demura@postman.riken.go.jp
            This clone was obtained at our laboratory
            Seq primer: M13 forward.
FEATURES             Location/Qualifiers
     source          1. .236
                     /organism="Zinnia elegans"
                     /mol_type="mRNA"
                     /cultivar="Canary bird"
                     /db_xref="taxon:34245"
                     /clone="Z10877"
                     /tissue_type="mesophyll cell"
                     /clone_lib="zinnia cultured mesophyll cell equalized cDNA"
                     /note="Vector: pGEM-T easy; cultured in tracheary element
                     differentiation-inductive medium"
ORIGIN

Alignment Scores:
Pred. No.:                 1.3e+03         Length:          236
Score:                     42.00           Matches:         6
Percent Similarity:        100.00%         Conservative:    0
Best Local Similarity:     100.00%         Mismatches:      0
Query Match:               100.00%         Indels:          0
DB:                        1               Gaps:            0

US-09-858-332G-28 (1-6) x AU302349 (1-236)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        235 CACAACCACAACCACAAC 218

RESULT 36
CO023831/c
LOCUS       CO023831                 237 bp    mRNA    linear   EST 10-JUN-2004
DEFINITION  EST788964 Coccidioides posadasii saprobic phase cDNA library, 2 to
            4 kb Coccidioides posadasii cDNA clone CIDCX01 5' end, mRNA
            sequence.
ACCESSION   CO023831
VERSION     CO023831.1  GI:48550268
KEYWORDS    EST.
SOURCE      Coccidioides posadasii
  ORGANISM  Coccidioides posadasii
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Onygenales; mitosporic Onygenales; Coccidioides.
REFERENCE   1  (bases 1 to 237)
  AUTHORS   Gardner,M.J. and Cole,G.T.
  TITLE     Analysis of gene expression in Coccidioides posadasii mycelia and
            spherules via expressed sequence tags
  JOURNAL   Unpublished (2003)
COMMENT     Other_ESTs: EST788963
            Contact: Gardner MJ
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301 838 3519
            Fax: 301 838 0208
            Email: gardner@tigr.org
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .237
                     /organism="Coccidioides posadasii"
                     /mol_type="mRNA"
                     /strain="C735"
                     /db_xref="taxon:199306"
                     /clone="CIDCX01"
                     /dev_stage="saprobic phase (mycelia)"
                     /lab_host="E. coli DH10B, T1 phage resistant"
                     /clone_lib="Coccidioides posadasii saprobic phase cDNA
                     library, 2 to 4 kb"
                     /note="Vector: pExpress 1; Site_1: Not I; Site_2: Eco RV;
                     Coccidioides posadasii saprobic phase cDNA library, size
                     fractionated cDNA 2 to 4 kb"
ORIGIN

Alignment Scores:
Pred. No.:                 1.31e+03        Length:          237
Score:                     42.00           Matches:         6
Percent Similarity:        100.00%         Conservative:    0
Best Local Similarity:     100.00%         Mismatches:      0
Query Match:               100.00%         Indels:          0
DB:                        7               Gaps:            0

US-09-858-332G-28 (1-6) x CO023831 (1-237)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        195 CACAACCACAACCACAAC 178

RESULT 37
AU073350
LOCUS       AU073350                 240 bp    mRNA    linear   EST 24-JUN-1999
DEFINITION  AU073350 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
            discoideum cDNA clone SSG748, mRNA sequence.
ACCESSION   AU073350
VERSION     AU073350.1  GI:5179771
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
```

```
REFERENCE   1  (bases 1 to 240)
  AUTHORS   Urushihara,H.
  TITLE     Developmental cDNA in Dictyostelium discoideum (1999)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp
            PROJECT = Dictyostelium discoideum cDNA project in Japan.
FEATURES             Location/Qualifiers
     source          1. .240
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SSG748"
                     /dev_stage="slug"
                     /clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
ORIGIN

Alignment Scores:
Pred. No.:              1.33e+03        Length:         240
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AU073350 (1-240)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        195 CACAACCACAACCACAAC 212

RESULT 38
CO476909/c
LOCUS       CO476909                  242 bp    mRNA    linear   EST 29-JUN-2005
DEFINITION  GQ0132.B3_I22 GQ013: Elongating ROOTS tips - saplings Picea glauca
            cDNA clone GenomeQuebec_Id:GQ0132I22 5', mRNA sequence.
ACCESSION   CO476909
VERSION     CO476909.1  GI:50150678
KEYWORDS    EST.
SOURCE      Picea glauca (white spruce)
  ORGANISM  Picea glauca
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
REFERENCE   1  (bases 1 to 242)
  AUTHORS   Morency,M.-J., Cooke,J., Pavy,N., Parsons,L., Paule,C., Seguin,A.,
            Retzel,E., Butterfield,Y., Barber,S., Yang,G., Stott,J.,
            Siddiqui,A., Holt,R., Marra,M. and MacKay,J.
  TITLE     Arborea EST sequencing in Picea glauca (white spruce)
  JOURNAL   Unpublished (2004)
COMMENT     Contact: John MacKay
            Centre de Recherche en Biologie Forestiere
            Universite Laval
            Pavillon Charles-Eugene Marchand, Quebec, Quebec, CANADA G1K 7P4
            Fax: 418 656 7493
            Email: jmackay@rsvs.ulaval.ca
            Center for Computational Genomics and Bioinformatics (CCGB),
            University of Minnesota, MN_id Identifier: MN5181462 Clone ID:
            GQ0132_I22 Clones available through: John MacKay, Ph. D. Professeur
            adjoint -Assistant professor EMAIL: jmackay@rsvs.ulaval.ca Centre
            de Recherche en Biologie Forestiere (Forest Biology Research
            Center) Universite Laval Quebec, Quebec CANADA G1K 7P4
            Plate: 2  row: 22  column: I
            Seq primer: T3 Primer.
FEATURES             Location/Qualifiers
     source          1. .242
                     /organism="Picea glauca"
                     /mol_type="mRNA"
                     /strain="pg-653"
                     /db_xref="taxon:3330"
                     /clone="GenomeQuebec_Id:GQ0132I22"
                     /sex="Hermaphrodite"
                     /tissue_type="Tender roots tips free of michorizee"
                     /dev_stage="Actively elongating roots"
                     /lab_host="E. coli DH10B cells"
                     /clone_lib="GQ013: Elongating ROOTS tips - saplings"
                     /note="Organ: Roots from 60 cm tall seedlings.; Vector:
                     pBluescript II SK (+) XR; Site_1: Eco-RI; Site_2: Xho-I;
                     cDNA was prepared from 5 mg of poly A+ selected RNA and
                     was directionally ligated into the pBluescript II SK (+)
                     XR vector (Stratagene), transformed by electroporation
                     into DH10B cells (In vitrogen) for propagation"
ORIGIN

Alignment Scores:
Pred. No.:              1.34e+03        Length:         242
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     7               Gaps:           0

US-09-858-332G-28 (1-6) x CO476909 (1-242)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         48 CACAATCACAACCACAAC 31

RESULT 39
AU303954/c
LOCUS       AU303954                  245 bp    mRNA    linear   EST 22-JAN-2004
DEFINITION  AU303954 zinnia cultured mesophyll cell equalized cDNA Zinnia
            elegans cDNA clone Z12935, mRNA sequence.
ACCESSION   AU303954
VERSION     AU303954.1  GI:41119878
KEYWORDS    EST.
SOURCE      Zinnia elegans
  ORGANISM  Zinnia elegans
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; campanulids; Asterales; Asteraceae; Asteroideae;
            Heliantheae; Zinnia.
REFERENCE   1  (bases 1 to 245)
  AUTHORS   Demura,T., Tashiro,G., Horiguchi,G., Kishimoto,N., Kubo,M.,
            Matsuoka,N., Minami,A., Nagata-Hiwatashi,M., Nakamura,K.,
            Okamura,Y., Sassa,N., Suzuki,S., Yazaki,J., Kikuchi,S. and
            Fukuda,H.
  TITLE     Visualization by comprehensive microarray analysis of gene
            expression programs during transdifferentiation of mesophyll cells
            into xylem cells
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (24), 15794-15799 (2002)
COMMENT     Contact: Taku Demura
            Morphogenesis Research Group
            RIKEN Plant Science Center
            1-7-22 Suehirocho, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9605
            Fax: 81-45-503-9573
            Email: demura@postman.riken.go.jp
            This clone was obtained at our laboratory
            Seq primer: M13 forward.
FEATURES             Location/Qualifiers
     source          1. .245
                     /organism="Zinnia elegans"
                     /mol_type="mRNA"
                     /cultivar="Canary bird"
                     /db_xref="taxon:34245"
                     /clone="Z12935"
                     /tissue_type="mesophyll cell"
                     /clone_lib="zinnia cultured mesophyll cell equalized cDNA"
                     /note="Vector: pGEM-T easy; cultured in tracheary element
                     differentiation-inductive medium"
```

```
ORIGIN

Alignment Scores:
Pred. No.:                1.35e+03        Length:         245
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       1               Gaps:           0

US-09-858-332G-28 (1-6) x AU303954 (1-245)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        128 CATAATCATAATCATAAT 111

RESULT 40
BU994168
LOCUS       BU994168                247 bp    mRNA    linear   EST 23-OCT-2002
DEFINITION  HM05P14r HM Hordeum vulgare subsp. vulgare cDNA clone HM05P14
            5-PRIME, mRNA sequence.
ACCESSION   BU994168
VERSION     BU994168.1  GI:24271151
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 247)
  AUTHORS   Zhang,H., Weschke,W., Michalek,W., Stein,N. and Graner,A.
  TITLE     EST sequencing and analysis in barley (2002)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Stein Nils
            Molecular Markers Group, Department Genbank
            Institute of Plant Genetics and Crop Plant Research (IPK)
            Correnssstr. 3, 06466, Gatersleben, Germany
            Tel: 039482-5522
            Fax: 039482-5595
            Email: stein@ipk-gatersleben.de
            Insert Length: 247   Std Error: 0.00
            Plate: 5  row: P  column: 14
            Seq primer: M13rev.
FEATURES             Location/Qualifiers
     source          1. .247
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Barke"
                     /sub_species="vulgare"
                     /db_xref="GABI:253954"
                     /db_xref="taxon:112509"
                     /clone="HM05P14"
                     /tissue_type="male inflorescences"
                     /dev_stage="male inflorescences (approx. 2 mm in size),
                     green anther stage"
                     /lab_host="XL10-Gold"
                     /clone_lib="HM"
                     /note="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of
                     cDNA); Site_2: XhoI (3'-end of cDNA); Due to a cloning
                     artefact caused by the kit, in most cases the EcoRI site
                     is NOT present, as well as the EcoRIadapter used for
                     cloning. To excise the insert, restriction sites upstream
                     EcoRI should be used (e.g. BamHI, SalI,PstI). NOTE: Also
                     due to the cloning system used Blue/white selection for
                     recombinats is not 100% reliable."
ORIGIN

Alignment Scores:
Pred. No.:                1.37e+03        Length:         247
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       5               Gaps:           0

US-09-858-332G-28 (1-6) x BU994168 (1-247)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        120 CACAACCACAACCACAAC 137

RESULT 41
CB606620
LOCUS       CB606620                250 bp    mRNA    linear   EST 16-MAY-2003
DEFINITION  AMGNNUC:CR1-00147-A9-B Colon Rat 1 (10390) Rattus norvegicus cDNA
            clone cr1-00147-a9 3', mRNA sequence.
ACCESSION   CB606620
VERSION     CB606620.1  GI:29546233
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 250)
  AUTHORS   Amgen EST Program.
  TITLE     Amgen Rat EST Program
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Dan Fitzpatrick
            Amgen, Inc
            One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA
            Tel: 805 447-4881
            Plate: 00147  row: a  column: 9.
FEATURES             Location/Qualifiers
     source          1. .250
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
                     /clone="cr1-00147-a9"
                     /tissue_type="intestine, fetal"
                     /clone_lib="Colon Rat 1 (10390)"
                     /note="Vector: pMOB; Site_1: SalI; Site_2: NotI; Colon Rat
                     1 Fetal rat intestinal cDNA library."
ORIGIN

Alignment Scores:
Pred. No.:                1.38e+03        Length:         250
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x CB606620 (1-250)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         84 CACAACCACAACCACAAC 101

RESULT 42
CW186051
LOCUS       CW186051                251 bp    DNA     linear   GSS 30-OCT-2004
DEFINITION  104_603_11166363_116_36701_002 Sorghum methylation filtered library
            (LibID: 104) Sorghum bicolor genomic clone 11166363, genomic survey
            sequence.
ACCESSION   CW186051
VERSION     CW186051.1  GI:54894379
KEYWORDS    GSS.
SOURCE      Sorghum bicolor (sorghum)
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 251)
  AUTHORS   Bedell,J.A., Budiman,M.A., Nunberg,A., Citek,R.W., Robbins,D.,
            Jones,J., Flick,E., Rohlfing,T., Fries,J., Bradford,K.,
```

```
            McMenamy,J., Smith,M., Holeman,H., Roe,B.A, Wiley,G., Korf,I.F.,
            Rabinowicz,P.D., Lakey,N., McCombie,W.R., Jeddeloh,J.A. and
            Martienssen,R.A.
  TITLE     Sorghum genome sequencing by methylation filtration
  JOURNAL   PLoS Biol. 3 (1), e13 (2005)
   PUBMED   15660154
COMMENT     Contact: Bedell JA
            Orion Genomics, LLC
            4041 Forest Park Ave, St. Louis, MO 63108, USA
            Tel: 314 615 6979
            Fax: 314 615 5975
            Email: jbedell@oriongenomics.com
            Plate: 603  row: p  column: 03
            Seq primer: T3 Reverse
            Class: methylation filtered
            High quality sequence stop: 251.
FEATURES             Location/Qualifiers
     source          1. .251
                     /organism="Sorghum bicolor"
                     /mol_type="genomic DNA"
                     /cultivar="ATx623"
                     /db_xref="taxon:4558"
                     /clone="11166363"
                     /clone_lib="Sorghum methylation filtered library (LibID:
                     104)"
                     /note="Organ: leaf; Vector: pBCSK(-); Site_1: HincII; DNA
                     prepared from purified nuclei was randomly sheared,
                     end-repaired, size fractionated to enrich for the 0.5 to 5
                     kb fraction, ligated into HincII-digested pBCSK(-) vector
                     and electroporated into E. coli cells. This is a
                     methylation filtered library."
ORIGIN

Alignment Scores:
Pred. No.:              1.39e+03        Length:         251
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     10              Gaps:           0

US-09-858-332G-28 (1-6) x CW186051 (1-251)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         91 CACAACCACAACCACAAC 108

RESULT 43
CC779608
LOCUS       CC779608                 253 bp    DNA     linear   GSS 01-JUL-2003
DEFINITION  ZMMBBc0431G01f ZMMBBc Zea mays genomic clone ZMMBBc0431G01 5',
            genomic survey sequence.
ACCESSION   CC779608
VERSION     CC779608.1  GI:32369382
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 253)
  AUTHORS   Bharti,A.K., Young,S., Kavchok,S., Keizer,G., Bronzino,A.C.,
            Rouzard,K., Fuks,G., Yu,Y., Wing,R. and Messing,J.
  TITLE     Sequencing of the maize genome at PGIR (2003b)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Bharti,A.K.
            Dr.Joachim Messing's lab
            The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers
            University
            190 Frelinghuysen Road, Piscataway, NJ 08854, USA
            Tel: 732 445 3801
            Fax: 732 445 5735
            Email: bharti@waksman.rutgers.edu
            Seq primer: T7
            Class: BAC ends
            High quality sequence start: 53.
FEATURES             Location/Qualifiers
     source          1. .253
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBBc0431G01"
                     /lab_host="E. coli DH10B"
                     /clone_lib="ZMMBBc"
                     /note="Vector: pTARBAC1.3; Site_1: BamHI; Site_2: BamHI"
ORIGIN

Alignment Scores:
Pred. No.:              1.4e+03         Length:         253
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     9               Gaps:           0

US-09-858-332G-28 (1-6) x CC779608 (1-253)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        187 CATAATCATAATCACAAT 204

RESULT 44
CG895908/c
LOCUS       CG895908                 253 bp    DNA     linear   GSS 08-DEC-2003
DEFINITION  pastbac005xf03.b1.ab1 Res147 1 Pasteuria penetrans genomic 5',
            genomic survey sequence.
ACCESSION   CG895908
VERSION     CG895908.1  GI:39551403
KEYWORDS    GSS.
SOURCE      Pasteuria penetrans
  ORGANISM  Pasteuria penetrans
            Bacteria; Firmicutes; Bacillales; Alicyclobacillaceae; Pasteuria.
REFERENCE   1  (bases 1 to 253)
  AUTHORS   Opperman,C.H., Davies,K.G., Sosinski,B.R., Waterman,J. and Burke,M.
  TITLE     Unpublished Data
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Opperman CH
            Center for the Biology of Nematode Parasitism and Nematode
            Interactions Unit
            North Carolina State University and Rothamsted Research, Ltd.
            Box 7253, NCSU, Raleigh, NC 27606, USA
            Email: warthog@unity.ncsu.edu
            Homology: e-val = 1e-10. Description: None found.  row: f  column:
            3
            Class: shotgun
            High quality sequence stop: 253.
FEATURES             Location/Qualifiers
     source          1. .253
                     /organism="Pasteuria penetrans"
                     /mol_type="genomic DNA"
                     /strain="Res147"
                     /db_xref="taxon:86005"
                     /clone_lib="Res147 1"
ORIGIN

Alignment Scores:
Pred. No.:              1.4e+03         Length:         253
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     10              Gaps:           0

US-09-858-332G-28 (1-6) x CG895908 (1-253)
```

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        251 CACAACCACAACCACAAC 234

RESULT 45
CB178412/c
LOCUS       CB178412                254 bp    mRNA    linear   EST 31-JAN-2003
DEFINITION  is38c09.y1 HR85 islet Homo sapiens cDNA clone IMAGE:6554682 5',
            mRNA sequence.
ACCESSION   CB178412
VERSION     CB178412.1  GI:28186802
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 254)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J.,
            Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R.,
            Williams,T., Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
COMMENT     Other_ESTs: is38c09.x1
            Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Dr. Hiroshi Inoue
            (hinoue@im.wustl.edu)
            Putative full length read
            vector to vector length is 255
            Seq primer: -40UP from Gibco.
FEATURES              Location/Qualifiers
     source           1. .254
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:6554682"
                      /tissue_type="Purified pancreatic islet"
                      /lab_host="DH10B"
                      /clone_lib="HR85 islet"
                      /note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
                      NotI; Site_2: XhoI; cDNA made by oligo-dT priming.
                      Size-selected on agarose gel. Average insert size ~1kb. 5'
                      XhoI site was destroyed after directional cloning.
                      Amplified once. Contact information: Hiroshi Inoue, MD,
                      Metabolism Div. (Alan Permutt Lab), Washington University
                      School of Medicine, Box 8127, 660 South Euclid Ave., St.
                      Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
                      314-362-1916, Fax: 314-747-2692."
ORIGIN

Alignment Scores:
Pred. No.:                1.41e+03        Length:          254
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       6               Gaps:            0

US-09-858-332G-28 (1-6) x CB178412 (1-254)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         41 CATAATCATAATCATAAT 24

RESULT 46
Z40263
LOCUS       Z40263                  254 bp    mRNA    linear   EST 07-NOV-1994
DEFINITION  HSC1XE012 normalized infant brain cDNA Homo sapiens cDNA clone
            c-1xe01 3', mRNA sequence.
ACCESSION   Z40263
VERSION     Z40263.1  GI:564879
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 254)
  AUTHORS   Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C.,
            Devignes,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B.,
            Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
            Sebastiani-Kabaktchis,C. and Tessier,A.
  TITLE     IMAGE: molecular integration of the analysis of the human genome
            and its expression
  JOURNAL   C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
  PUBMED    7757816
COMMENT     Contact: Genethon
            Genexpress-Genethon
            Genethon Centre de recherche sur le Genome Humain
            1,rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
            Tel: 33169472800
            Fax: 33160778698
            Email: genexpress@genethon.fr
            Single read. 18 T removed at sequence 5'end
            Genexpress_library_idt: C; Genexpress_sequence_idt: a1c-1xe01
            Seq primer: (-21)M13_universal.
FEATURES              Location/Qualifiers
     source           1. .254
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="c-1xe01"
                      /sex="Female"
                      /tissue_type="total brain"
                      /dev_stage="3 months old"
                      /clone_lib="normalized infant brain cDNA"
                      /note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
                      Site_2: NotI; sex=Female; dev_stage=3 months old;
                      isolate=muscular atrophy patient;  tissue_type=total
                      brain;  total mRNA was oligo-(dT) primed and directionally
                      cloned 5' -> 3' into the  HindIII -> NotI sites of the
                      lafmid BA vector. Clone library from B.Soares,  Psychiatry
                      Dept. Columbia University, USA.  Normalization_method:
                      Bento Soares, P.N.A.S in press"
ORIGIN

Alignment Scores:
Pred. No.:                1.41e+03        Length:          254
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       8               Gaps:            0

US-09-858-332G-28 (1-6) x Z40263 (1-254)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        191 CATAATCATAATCATAAT 208

RESULT 47
BM522273
LOCUS       BM522273                258 bp    mRNA    linear   EST 19-FEB-2002
DEFINITION  ESSU0811 S.scabiei cDNA library Sarcoptes scabiei cDNA clone
```

```
            SAS0948 5', mRNA sequence.
ACCESSION   BM522273
VERSION     BM522273.1  GI:18707309
KEYWORDS    EST.
SOURCE      Sarcoptes scabiei
  ORGANISM  Sarcoptes scabiei
            Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
            Acariformes; Sarcoptiformes; Astigmata; Psoroptidia; Sarcoptoidea;
            Sarcoptidae; Sarcoptes.
REFERENCE   1  (bases 1 to 258)
  AUTHORS   Ljunggren,E.L., Nilsson,D., Naslund,K. and Mattsson,J.G.
  TITLE     Expressed sequence tag analysis of the parasitic mite Sarcoptes
            scabiei
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Mattsson J.G.
            Department of Parasitology (SWEPAR)
            National Veterinary Institute
            SE-751 89 Uppsala, Sweden
            Tel: +46 18 674120
            Fax: +46 18
            Email: jens.mattsson@sva.se
            Seq primer: T3 primer
            High quality sequence stop: 258.
FEATURES             Location/Qualifiers
     source          1. .258
                     /organism="Sarcoptes scabiei"
                     /mol_type="mRNA"
                     /db_xref="taxon:52283"
                     /clone="SAS0948"
                     /clone_lib="S.scabiei cDNA library"
                     /note="The Sarcoptes scabiei mixed lifestage library was
                     constructed by Jens G Mattsson. cDNAs were synthesized
                     from poly(A)+ RNA by oligo d(T) priming, size-selected and
                     directionally cloned into the Uni-ZAP lambda vector
                     (Stratagene). The primary library was amplified on
                     XL1-Blue MRF' cells."
ORIGIN

Alignment Scores:
Pred. No.:              1.43e+03        Length:         258
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-858-332G-28 (1-6) x BM522273 (1-258)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        124 CATAACCATAATCATAAC 141

RESULT 48
AJ497920/c
LOCUS       AJ497920                262 bp    mRNA    linear   EST 09-AUG-2002
DEFINITION  AJ497920 MTFLOW Medicago truncatula cDNA clone mt--abc955116e03,
            mRNA sequence.
ACCESSION   AJ497920
VERSION     AJ497920.1  GI:22088363
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 262)
  AUTHORS   Firnhaber,C., Bartelsmeier,V., Becker,A., Meyer,F., Bartels,D.,
            Bekel,T., Linke,B., Puehler,A. and Kuester,H.
  TITLE     Determination of transcript sequences from developing flowers of
            Medicago truncatula genotype A17
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Kuester H
            Lehrstuhl fuer Genetik
            Universitaet Bielefeld
            Postfach 100131, D-33501 Bielefeld, Germany.
FEATURES             Location/Qualifiers
     source          1. .262
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /db_xref="taxon:3880"
                     /clone="mt--abc955116e03"
                     /tissue_type="flowers"
                     /dev_stage="different stages of development"
                     /clone_lib="MTFLOW"
                     /note="Vector: pGEM-T; Site_1: T-overhang; Site_2: XhoI;
                     genotype A17; cDNA was prepared from polyA+ enriched RNA
                     from flowers harvested at different stages of development.
                     The cDNA was directionally ligated into the pGEM-T vector
                     from Promega. Plasmids containing cDNA inserts were
                     propagated in E. coli DH5alpha cells."
ORIGIN

Alignment Scores:
Pred. No.:              1.46e+03        Length:         262
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AJ497920 (1-262)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         57 CACAATCATAATCATAAT 40

RESULT 49
CZ918340
LOCUS       CZ918340                262 bp    DNA     linear   GSS 08-AUG-2005
DEFINITION  4021009A04.2EL_x1 4021 - RescueMu Grid V Zea mays genomic, genomic
            survey sequence.
ACCESSION   CZ918340
VERSION     CZ918340.1  GI:71937774
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 262)
  AUTHORS   Walbot,V.
  TITLE     Maize genomic sequences found using engineered RescueMu transposon
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Possible ligation site of ends cut by 2 different endonucleases.
            Reverse complemented post-ligation sequence from source sequence.
            Plate: 4021009  row: 1
            Class: transposon-tagged.
FEATURES             Location/Qualifiers
     source          1. .262
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="mixed background W23/A188/B73/K55"
                     /db_xref="taxon:4577"
                     /tissue_type="leaf"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="4021 - RescueMu Grid V"
                     /note="Organ: leaf; Vector: RescueMu (engineered from
```

pBlueScript backbone); Site_1: BamHI; Site_2: BglII; RescueMu is a 4.9 kb, modified maize Mu transposon designed to allow plasmid rescue from total genomic DNA. Mu elements insert preferentially into transcription units. For more information on RescueMu, go to the web site 'http://www.mutransposon.org/project/RescueMu/'. Grid V was grown at University of Arizona in 2003. DNA was extracted from leaf strips, double digested using BamHI and BglII, and ligated to form circular plasmids. DH10B cells were transformed and then screened on LB plates with ampicillin."

ORIGIN

Alignment Scores:

Pred. No.:	1.46e+03	Length:	262
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	10	Gaps:	0

US-09-858-332G-28 (1-6) x CZ918340 (1-262)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        223 CATAACCATAACCATAAC 240
```

RESULT 50
AU053106
LOCUS AU053106 264 bp mRNA linear EST 28-APR-1999
DEFINITION AU053106 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium discoideum cDNA clone SLF703, mRNA sequence.
ACCESSION AU053106
VERSION AU053106.1 GI:4701589
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 264)
AUTHORS Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M., Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H., Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
TITLE Developmental cDNA in Dictyostelium discoideum
JOURNAL Unpublished (1998)
COMMENT Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
Tel: 81-298-53-4664
Fax: 81-298-53-6614
Email: hideko@biol.tsukuba.ac.jp
PROJECT = Dictyostelium discoideum cDNA project in Japan.
FEATURES Location/Qualifiers
source 1. .264
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="SLF703"
/dev_stage="slug"
/clone_lib="Dictyostelium discoideum SL (H.Urushihara)"

ORIGIN

Alignment Scores:

Pred. No.:	1.47e+03	Length:	264
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0

US-09-858-332G-28 (1-6) x AU053106 (1-264)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         41 CACAATCACAATCACAAT 58
```

RESULT 51
BF428793
LOCUS BF428793 269 bp mRNA linear EST 29-NOV-2000
DEFINITION WHE1711_E08_J15ZS Wheat heat stressed spike cDNA library Triticum aestivum cDNA clone WHE1711_E08_J15, mRNA sequence.
ACCESSION BF428793
VERSION BF428793.1 GI:11440714
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 269)
AUTHORS Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Malatrasi,M., Miller,R., Nguyen,H.T., Rausch,C.J., Seaton,C.L., Tong,J.C. and Zhang,D.
TITLE The structure and function of the expressed portion of the wheat genomes - Heat stressed spike cDNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oandersn@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
Seq primer: Stratagene SK primer.
FEATURES Location/Qualifiers
source 1. .269
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE1711_E08_J15"
/tissue_type="Whole spike"
/dev_stage="Spikes at 5, 10, 15 and 20 days after anthesis"
/lab_host="E. coli SOLR"
/clone_lib="Wheat heat stressed spike cDNA library"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site_1: EcoRI; Site_2: XhoI; Spikes at 5, 10, 15 and 20 days after anthesis were heat stressed under two conditions at Texas Tech University (D. Zhang in HT Nguyrn lab): (1) at 38 C for 4 hours and (2) 5 days of cyclic treatment of 38 C for 4 hours. Total RNA and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give pBluescript phagemids in the TJ Close lab (Choi, Close, Fenton, Malatrasi) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

ORIGIN

Alignment Scores:

Pred. No.:	1.5e+03	Length:	269
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-28 (1-6) x BF428793 (1-269)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        165 CACAACCACAACCACAAT 182
```

```
RESULT 52
BE759303/c
LOCUS       BE759303                269 bp    mRNA    linear   EST 24-OCT-2000
DEFINITION  an_3656 Aspergillus niger, pYES2 (XhoI-EcoRI) Aspergillus niger
            cDNA clone 3656 5', mRNA sequence.
ACCESSION   BE759303
VERSION     BE759303.1  GI:10181940
KEYWORDS    EST.
SOURCE      Aspergillus niger
  ORGANISM  Aspergillus niger
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE   1  (bases 1 to 269)
  AUTHORS   Tsang,A. and Storms,R.
  TITLE     Aspergillus niger Expressed Sequence Tags
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Tsang, A.
            Dept. of Biology
            Concordia University
            1455 deMaisonneuve Blvd. West, Montreal, Quebec, CANADA, H3G 1M8
            Tel: 514-848-3405
            Fax: 514-848-4504
            Email: tsang@vax2.concordia.ca
            PCR PRimers
            FORWARD: CGACATTATGCTGAGTGATATCCC
            Insert Length: 269   Std Error: 0.00
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .269
                     /organism="Aspergillus niger"
                     /mol_type="mRNA"
                     /db_xref="taxon:5061"
                     /clone="3656"
                     /lab_host="E. coli"
                     /clone_lib="Aspergillus niger, pYES2 (XhoI-EcoRI)"
                     /note="Vector: pYES2; Site_1: Xho-I; Site_2: EcoRI; cDNA
                     was synthesized with ZAP kit (Stratagene) using poly(A)+
                     RNA isolated from Aspergillus niger grown in both complete
                     and minimal media. Synthesis was primed with oligo(dT)
                     primer/XhoI-linker. EcoRI adaptors were later ligated to
                     polished ends. EcoRI-XhoI-digested cDNA was ligated with
                     EcoRI-XhoI-digested pYES2 (Invitrogen Corp). This vector
                     permits expression of our library in yeast. "
ORIGIN

Alignment Scores:
Pred. No.:              1.5e+03       Length:         269
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     2             Gaps:           0

US-09-858-332G-28 (1-6) x BE759303 (1-269)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        101 CATAACCATAACCATAAC 84

RESULT 53
CE679130/c
LOCUS       CE679130                271 bp    DNA     linear   GSS 29-SEP-2003
DEFINITION  tigr-gss-dog-17000313989364 Dog Library Canis familiaris genomic,
            genomic survey sequence.
ACCESSION   CE679130
VERSION     CE679130.1  GI:36998130
KEYWORDS    GSS.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
            Canis.
REFERENCE   1  (bases 1 to 271)
  AUTHORS   Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
            Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
            Venter,J.C.
  TITLE     The dog genome: survey sequencing and comparative analysis
  JOURNAL   Science 301 (5641), 1898-1903 (2003)
   PUBMED   14512627
COMMENT     Contact: Kirkness EF
            The Institute for Genomic Research
            Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
            Rockville, MD 20850, USA
            Tel: 301-838-0200
            Fax: 301-838-0208
            Email: ekirknes@tigr.org
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .271
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /strain="Standard Poodle"
                     /db_xref="taxon:9615"
                     /clone_lib="Dog Library"
                     /note="Site_1: BstXI; Libraries were prepared from
                     peripheral blood"
ORIGIN

Alignment Scores:
Pred. No.:              1.51e+03      Length:         271
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     10            Gaps:           0

US-09-858-332G-28 (1-6) x CE679130 (1-271)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         96 CACAACCACAACCATAAC 79

RESULT 54
CB178653
LOCUS       CB178653                275 bp    mRNA    linear   EST 31-JAN-2003
DEFINITION  is38c09.x1 HR85 islet Homo sapiens cDNA clone IMAGE:6554682 3',
            mRNA sequence.
ACCESSION   CB178653
VERSION     CB178653.1  GI:28187043
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 275)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J.,
            Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R.,
            Williams,T., Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
COMMENT     Other_ESTs: is38c09.y1
            Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
            Washington University Genome Sequencing Center For information on
```

obtaining a clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
Seq primer: -40RP from Gibco.

```
FEATURES             Location/Qualifiers
     source          1. .275
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6554682"
                     /tissue_type="Purified pancreatic islet"
                     /lab_host="DH10B"
                     /clone_lib="HR85 islet"
                     /note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
                     NotI; Site_2: XhoI; cDNA made by oligo-dT priming.
                     Size-selected on agarose gel. Average insert size ~1kb. 5'
                     XhoI site was destroyed after directional cloning.
                     Amplified once. Contact information: Hiroshi Inoue, MD,
                     Metabolism Div. (Alan Permutt Lab), Washington University
                     School of Medicine, Box 8127, 660 South Euclid Ave., St.
                     Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
                     314-362-1916, Fax: 314-747-2692."
ORIGIN

Alignment Scores:
Pred. No.:                1.53e+03       Length:          275
Score:                    42.00          Matches:         6
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       6              Gaps:            0

US-09-858-332G-28 (1-6) x CB178653 (1-275)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        203 CATAATCATAATCATAAT 220

RESULT 55
BH613103
LOCUS       BH613103                276 bp    DNA     linear   GSS 04-JAN-2002
DEFINITION  SALK_033777 Arabidopsis thaliana TDNA insertion lines Arabidopsis
            thaliana genomic clone SALK_033777, genomic survey sequence.
ACCESSION   BH613103
VERSION     BH613103.1  GI:18061762
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 276)
  AUTHORS   Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,
            Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,
            Shinn,P., Zimmerman,J. and Ecker,J.R.
  TITLE     A Sequence-Indexed Library of Insertion Mutations in the
            Arabidopsis Genome
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Joseph R. Ecker
            Salk Institute Genomic Analysis Laboratory (SIGnAL)
            The Salk Institute for Biological Studies
            10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
            Tel: 858 453 4100 x1752
            Fax: 858 558 6379
            Email: ecker@salk.edu
            This is single pass sequence recovered from the left border of
            TDNA.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1. .276
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /ecotype="Col-0"
                     /db_xref="taxon:3702"
                     /clone="SALK_033777"
                     /clone_lib="Arabidopsis thaliana TDNA insertion lines"
                     /note="PCR was performed on Arabidopsis thaliana lines
                     each of which contains  one or more TDNA insertion
                     elements. The resultant fragment for each  line was
                     directly sequenced to determine the genomic sequence at
                     the  site of insertion. Details of the protocols used can
                     be found at  http://signal.salk.edu/tdna_protocols.html"
ORIGIN

Alignment Scores:
Pred. No.:                1.54e+03       Length:          276
Score:                    42.00          Matches:         6
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       9              Gaps:            0

US-09-858-332G-28 (1-6) x BH613103 (1-276)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        101 CATAATCATAACCATAAC 118

RESULT 56
BG459579
LOCUS       BG459579                277 bp    mRNA    linear   EST 19-MAR-2001
DEFINITION  M77022STM Arabidopsis developing seed Arabidopsis thaliana cDNA
            clone 600039888R1 5', mRNA sequence.
ACCESSION   BG459579
VERSION     BG459579.1  GI:13382904
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 277)
  AUTHORS   White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de
            Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.
  TITLE     A new set of Arabidopsis expressed sequence tags from developing
            seeds. The metabolic pathway from carbohydrates to seed oil
  JOURNAL   Plant Physiol. 124 (4), 1582-1594 (2000)
   PUBMED   11115876
COMMENT     Contact: Benning, C
            Dept. of Biochemistry & Molecular Biology
            Michigan State University
            224 Biochemistry, Michigan State University, East Lansing, MI
            48824, USA
            Tel: 517 355 1609
            Fax: 517 353 9334
            Email: benning@msu.edu
            Clones were originally prepared at Michigan State University.
            Arabidopsis Biological Resource Center, The Ohio State University,
            309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210
            USA, FAX: 6142920603 TEL: 6142929371.
FEATURES             Location/Qualifiers
     source          1. .277
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /ecotype="Columbia"
                     /db_xref="taxon:3702"
                     /clone="600039888R1"
                     /tissue_type="seed"
                     /dev_stage="5-13 days after flowering"
                     /lab_host="E.coli"
                     /clone_lib="Arabidopsis developing seed"
                     /note="Organ: Developing seed; Vector: pBluescript SK-;
                     Site_1: EcoRI; Site_2: XhoII"
ORIGIN

Alignment Scores:
Pred. No.:                1.54e+03       Length:          277
```

```
Score:                 42.00         Matches:       6
Percent Similarity:    100.00%       Conservative:  0
Best Local Similarity: 100.00%       Mismatches:    0
Query Match:           100.00%       Indels:        0
DB:                    2             Gaps:          0

US-09-858-332G-28 (1-6) x BG459579 (1-277)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db          3 CACAATCATAATCACAAT 20

RESULT 57
H46261
LOCUS       H46261                   279 bp    mRNA    linear   EST 31-JUL-1995
DEFINITION  yo15d06.s1 Soares adult brain N2b5HB55Y Homo sapiens cDNA clone
            IMAGE:177995 3', mRNA sequence.
ACCESSION   H46261
VERSION     H46261.1  GI:922313
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 279)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1010
            High quality sequence stops: 209
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1010   Std Error: 0.00
            Seq primer: SP6
            High quality sequence stop: 209.
FEATURES             Location/Qualifiers
     source          1. .279
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="GDB:3840191"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:177995"
                     /sex="Male"
                     /dev_stage="55-year old"
                     /lab_host="DH10B (ampicillin resistant)"
                     /clone_lib="Soares adult brain N2b5HB55Y"
                     /note="Organ: brain; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 53. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. The adult brain RNA was
                     provided by Dr. Donald H. Gilden. Tissue was acquired
                     17-18 hours after death which occurred in consequence of a
                     ruptured aortic aneurysm. RNA was prepared from a pool of
                     tissues representing the following areas of the brain:
                     frontal, parietal, temporal and occipital cortex from the
                     left and right hemispheres, subcortical white matter,
                     basal ganglia, thalamus, cerebellum, midbrain, pons and
                     medulla."
ORIGIN

Alignment Scores:
Pred. No.:             1.56e+03      Length:        279
Score:                 42.00         Matches:       6
Percent Similarity:    100.00%       Conservative:  0
Best Local Similarity: 100.00%       Mismatches:    0
Query Match:           100.00%       Indels:        0
DB:                    8             Gaps:          0

US-09-858-332G-28 (1-6) x H46261 (1-279)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        177 CATAATCATAATCATAAT 194

RESULT 58
CA264926/c
LOCUS       CA264926                 281 bp    mRNA    linear   EST 26-SEP-2003
DEFINITION  SCAGRT3048D03.g RT3 Saccharum officinarum cDNA clone SCAGRT3048D03
            5', mRNA sequence.
ACCESSION   CA264926
VERSION     CA264926.1  GI:35960936
KEYWORDS    EST.
SOURCE      Saccharum officinarum
  ORGANISM  Saccharum officinarum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Saccharum; Saccharum officinarum
            complex.
REFERENCE   1  (bases 1 to 281)
  AUTHORS   Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
  TITLE     The libraries that made SUCEST
  JOURNAL   Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
COMMENT     Contact: Arruda P
            Centro de Biologia Molecular e Engenharia Genetica
            Universidade Estadual de Campinas
            Caixa Postal 6010, 13083-970, Campinas SP, Brazil
            Tel: 55 19 3788 1137
            Fax: 55 19 3788 1089
            Email: parruda@unicamp.br
            Clone distribution: clone distribution information can be found
            through the Brazilian Clone Collection Center (BCCC) at
            http://www.bcccenter.fcav.unesp.br
            Plate: 048  row: D  column: 03
            Seq primer: T7 Promoter Primer.
FEATURES             Location/Qualifiers
     source          1. .281
                     /organism="Saccharum officinarum"
                     /mol_type="mRNA"
                     /db_xref="taxon:4547"
                     /clone="SCAGRT3048D03"
                     /lab_host="DH10B"
                     /clone_lib="RT3"
                     /note="Organ: Root apex from adult plants; Vector:
                     pSport1; Site_1: SalI; Site_2: NotI; An unidirectional
                     cDNA library generated from [Root apex from adult
                     plants]. cDNA was prepared from polyA+ mRNA using
                     SuperScript Plasmid  System Kit (Invitrogen). The
                     double-strand cDNAs were fractionated in a  sepharose
                     CL-2B 40cm-columns and fragments sizing between 0.8 and
                     1.5 Kb  were directionally cloned into the vector. Details
                     of each source of RNA  and library construction can be
                     obtained at  http://sucest.lad.ic.unicamp.br/public"
ORIGIN

Alignment Scores:
Pred. No.:             1.57e+03      Length:        281
Score:                 42.00         Matches:       6
Percent Similarity:    100.00%       Conservative:  0
```

```
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     6              Gaps:           0

US-09-858-332G-28 (1-6) x CA264926 (1-281)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        213 CACAACCACAACCACAAC 196

RESULT 59
W40684
LOCUS       W40684                   285 bp    mRNA    linear   EST 20-MAY-1996
DEFINITION  mc66a06.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:353458 5' similar to gb:M38188 OVARIAN GRANULOSA CELL
            13.0 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION   W40684
VERSION     W40684.1  GI:1325183
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 285)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:225258
            Seq primer: mob.REGA+ET
            High quality sequence stop: 216.
FEATURES             Location/Qualifiers
     source          1. .285
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:353458"
                     /sex="unknown"
                     /tissue_type="embryo"
                     /dev_stage="13.5-14.5dpc total fetus"
                     /lab_host="DH10B"
                     /clone_lib="Soares mouse embryo NbME13.5 14.5"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTTT
                     T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                     14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                     State Univ., from 2 ]; double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector. Library went through one round of
                     normalization, and was constructed by Bento Soares and
                     M.Fatima Bonaldo.  "
ORIGIN

Alignment Scores:
Pred. No.:              1.59e+03       Length:         285
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     8              Gaps:           0

US-09-858-332G-28 (1-6) x W40684 (1-285)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        244 CACAACCATAACCACAAC 261

RESULT 60
BE759955/c
LOCUS       BE759955                 286 bp    mRNA    linear   EST 24-OCT-2000
DEFINITION  an_2902 Aspergillus niger, pYES2 (XhoI-EcoRI) Aspergillus niger
            cDNA clone 2902 5', mRNA sequence.
ACCESSION   BE759955
VERSION     BE759955.1  GI:10182592
KEYWORDS    EST.
SOURCE      Aspergillus niger
  ORGANISM  Aspergillus niger
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE   1  (bases 1 to 286)
  AUTHORS   Tsang,A. and Storms,R.
  TITLE     Aspergillus niger Expressed Sequence Tags
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Tsang, A.
            Dept. of Biology
            Concordia University
            1455 deMaisonneuve Blvd. West, Montreal, Quebec, CANADA, H3G 1M8
            Tel: 514-848-3405
            Fax: 514-848-4504
            Email: tsang@vax2.concordia.ca
            PCR PRimers
            FORWARD: CGACATTATGCTGAGTGATATCCC
            Insert Length: 286   Std Error: 0.00
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .286
                     /organism="Aspergillus niger"
                     /mol_type="mRNA"
                     /db_xref="taxon:5061"
                     /clone="2902"
                     /lab_host="E. coli"
                     /clone_lib="Aspergillus niger, pYES2 (XhoI-EcoRI)"
                     /note="Vector: pYES2; Site_1: Xho-I; Site_2: EcoRI; cDNA
                     was synthesized with ZAP kit (Stratagene) using poly(A)+
                     RNA isolated from Aspergillus niger grown in both complete
                     and minimal media. Synthesis was primed with oligo(dT)
                     primer/XhoI-linker. EcoRI adaptors were later ligated to
                     polished ends. EcoRI-XhoI-digested cDNA was ligated with
                     EcoRI-XhoI-digested pYES2 (Invitrogen Corp). This vector
                     permits expression of our library in yeast. "
ORIGIN

Alignment Scores:
Pred. No.:              1.6e+03        Length:         286
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     2              Gaps:           0

US-09-858-332G-28 (1-6) x BE759955 (1-286)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        200 CACAATCACAATCACAAT 183

RESULT 61
AV419462
LOCUS       AV419462                 287 bp    mRNA    linear   EST 10-MAR-2004
```

```
DEFINITION  AV419462 Lotus japonicus young plants (two-week old) Lotus
            corniculatus var. japonicus cDNA clone MWM169c09_r 5', mRNA
            sequence.
ACCESSION   AV419462
VERSION     AV419462.1  GI:7748641
KEYWORDS    EST.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1  (bases 1 to 287)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Generation of 7137 non-redundant expressed sequence tags from a
            legume, Lotus japonicus
  JOURNAL   DNA Res. 7 (2), 127-130 (2000)
   PUBMED   10819328
COMMENT     Contact: Yasukazu Nakamura
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: ynakamu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .287
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="mRNA"
                     /isolate="Miyakojima MG-20"
                     /db_xref="taxon:34305"
                     /clone="MWM169c09_r"
                     /tissue_type="whole plant"
                     /dev_stage="two-week old plant"
                     /clone_lib="Lotus japonicus young plants (two-week old)"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI; synonym: Lotus japonicus"
ORIGIN

Alignment Scores:
Pred. No.:                1.6e+03        Length:          287
Score:                    42.00          Matches:         6
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       1              Gaps:            0

US-09-858-332G-28 (1-6) x AV419462 (1-287)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        144 CACAATCACAATCACAAT 161

RESULT 62
CO813747/c
LOCUS       CO813747                287 bp    mRNA    linear   EST 06-AUG-2004
DEFINITION  AGENCOURT_30336371 NIH_ZGC_8 Danio rerio cDNA clone IMAGE:7410739
            5', mRNA sequence.
ACCESSION   CO813747
VERSION     CO813747.1  GI:51032373
KEYWORDS    EST.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 287)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Takuya Sakaguchi and Didier Stainier
            cDNA Library Preparation: Dr. Sumio Sugano
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM15617  row: f  column: 17
            High quality sequence stop: 288.
FEATURES             Location/Qualifiers
     source          1. .287
                     /organism="Danio rerio"
                     /mol_type="mRNA"
                     /db_xref="taxon:7955"
                     /clone="IMAGE:7410739"
                     /lab_host="DH10B TonA"
                     /clone_lib="NIH_ZGC_8"
                     /note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII;
                     Site_2: DraIII"
ORIGIN

Alignment Scores:
Pred. No.:                1.6e+03        Length:          287
Score:                    42.00          Matches:         6
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       7 *            Gaps:            0

US-09-858-332G-28 (1-6) x CO813747 (1-287)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        273 CACAATCACAATCACAAT 256

RESULT 63
CL392184
LOCUS       CL392184                287 bp    DNA     linear   GSS 05-MAR-2004
DEFINITION  ZMMBBb0392H04f ZMMBBb (HindIII) Zea mays genomic clone
            ZMMBBb0392H04 5', genomic survey sequence.
ACCESSION   CL392184
VERSION     CL392184.1  GI:45188558
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 287)
  AUTHORS   Bharti,A.K., Young,S., Kavchok,S., Keizer,G., Bronzino,A.C.,
            Zohovetz,V., Fuks,G., Yu,Y., Wing,R. and Messing,J.
  TITLE     Sequencing of the maize genome at PGIR (2003c)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Bharti,A.K.
            Dr.Joachim Messing's lab
            The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers
            University
            190 Frelinghuysen Road, Piscataway, NJ 08854, USA
            Tel: 732 445 3801
            Fax: 732 445 5735
            Email: bharti@waksman.rutgers.edu
            Seq primer: T7
            Class: BAC ends
            High quality sequence start: 103.
FEATURES             Location/Qualifiers
     source          1. .287
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBBb0392H04"
                     /lab_host="E. coli DH10B"
                     /clone_lib="ZMMBBb (HindIII)"
```

```
                     /note="Vector: pCUGI; Site_1: HindIII; Site_2: HindIII"
ORIGIN

Alignment Scores:
Pred. No.:                1.6e+03        Length:          287
Score:                    42.00          Matches:         6
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       10             Gaps:            0

US-09-858-332G-28 (1-6) x CL392184 (1-287)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        164 CATAATCATAATCATAAT 181

RESULT 64
CN252326
LOCUS       CN252326                289 bp    mRNA    linear   EST 09-APR-2004
DEFINITION  EST018258 Mycelium and yeast cells from Paracoccidioides
            brasiliensis Paracoccidioides brasiliensis cDNA, mRNA sequence.
ACCESSION   CN252326
VERSION     CN252326.1  GI:46356072
KEYWORDS    EST.
SOURCE      Paracoccidioides brasiliensis
  ORGANISM  Paracoccidioides brasiliensis
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Onygenales; mitosporic Onygenales; Paracoccidioides.
REFERENCE   1  (bases 1 to 289)
  AUTHORS   Felipe,M.S., Andrade,R.V., Arraes,F.B., Nicola,A.M., Maranhao,A.Q.,
            Torres,F.A., Silva-Pereira,I., Pocas-Fonseca,M.J., Campos,E.G.,
            Moraes,L.M., Andrade,P.A., Tavares,A.H., Silva,S.S., Kyaw,C.M.,
            Souza,D.P., Network,P., Pereira,M., Jesuino,R.S., Andrade,E.V.,
            Parente,J.A., Oliveira,G.S., Barbosa,M.S., Martins,N.F.,
            Fachin,A.L., Cardoso,R.S., Passos,G.A., Almeida,N.F., Walter,M.E.,
            Soares,C.M., Carvalho,M.J. and Brigido,M.M.
  TITLE     Transcriptional Profiles of the Human Pathogenic Fungus
            Paracoccidioides brasiliensis in Mycelium and Yeast Cells
  JOURNAL   J. Biol. Chem. 280 (26), 24706-24714 (2005)
   PUBMED   15849188
COMMENT     Contact: Felipe MSS
            Laboratory of Molecular Biology
            Institute of Biology - University of Brasilia
            Campus Universitario, Asa Norte, Brasilia, DF 70910-900, BRA
            Tel: 55 61 307 2423
            Fax: 55 61 349 8411
            Email: msueli@unb.br
            Seq primer: T7 Sequencing primer.
FEATURES             Location/Qualifiers
     source          1. .289
                     /organism="Paracoccidioides brasiliensis"
                     /mol_type="mRNA"
                     /strain="Pb01"
                     /db_xref="taxon:121759"
                     /clone_lib="Mycelium and yeast cells from Paracoccidioides
                     brasiliensis"
                     /note="Pb Lambda Zap Express Library"
ORIGIN

Alignment Scores:
Pred. No.:                1.62e+03       Length:          289
Score:                    42.00          Matches:         6
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       7              Gaps:            0

US-09-858-332G-28 (1-6) x CN252326 (1-289)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         84 CATAACCACAACCACAAC 101

RESULT 65
BQ986559/c
LOCUS       BQ986559                290 bp    mRNA    linear   EST 21-AUG-2002
DEFINITION  QGE9N08.yg.ab1 QG_EFGHJ lettuce serriola Lactuca sativa cDNA clone
            QGE9N08, mRNA sequence.
ACCESSION   BQ986559
VERSION     BQ986559.1  GI:22404084
KEYWORDS    EST.
SOURCE      Lactuca sativa
  ORGANISM  Lactuca sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
            Cichorieae; Lactuca.
REFERENCE   1  (bases 1 to 290)
  AUTHORS   Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J.,
            Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y.,
            Lai,Z., Church,S., Jackson,L. and Bradford,K.
  TITLE     Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
            belongs to contig QG_CA_Contig3631, see http://cgpdb.ucdavis.edu/
            for details.
            Plate: QGE9  row: N  column: 08.
FEATURES             Location/Qualifiers
     source          1. .290
                     /organism="Lactuca sativa"
                     /mol_type="mRNA"
                     /cultivar="L.serriola"
                     /db_xref="taxon:4236"
                     /clone="QGE9N08"
                     /lab_host="E.coli"
                     /clone_lib="QG_EFGHJ lettuce serriola"
                     /note="Vector: pBRcDNASfiAB; The library was constructed
                     from 10 different sources of RNA from a single genotype.
                     Separate cDNAs were generated using primers that
                     incorporated unique 5' and 3' tags to distinguish each
                     source of RNA. cDNAs were then pooled, size-fractionated,
                     directionally cloned into a custom medium-copy vector and
                     transformations made with four size classes to minimize
                     size bias. Details of each source of RNA and library
                     construction can be obtained at http://cgpdb.ucdavis.edu/
                     TAG_TISSUE=flowers environmental stress
                     TAG_LIB=QG_EFGHJ lettuce serriola
                     TAG_SEQ=CGAATGCGGG"
ORIGIN

Alignment Scores:
Pred. No.:                1.62e+03       Length:          290
Score:                    42.00          Matches:         6
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       5              Gaps:            0

US-09-858-332G-28 (1-6) x BQ986559 (1-290)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        197 CATAATCATAATCATAAT 180

RESULT 66
CF986668
LOCUS       CF986668                290 bp    mRNA    linear   EST 04-FEB-2005
```

```
DEFINITION  24212rsicee_5870.y1 Oryza sativa cv. LYP9 tillering whole plant
            cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF986668
VERSION     CF986668.1  GI:58588360
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 290)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 290
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .290
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="tillering"
                     /clone_lib="Oryza sativa cv. LYP9 tillering whole plant
                     cDNA library"
ORIGIN

Alignment Scores:
Pred. No.:              1.62e+03        Length:         290
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     7               Gaps:           0

US-09-858-332G-28 (1-6) x CF986668 (1-290)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        213 CACAACCACAACCACAAC 230

RESULT 67
AZ278916/c
LOCUS       AZ278916                291 bp    DNA     linear   GSS 26-JUL-2000
DEFINITION  RPCI-23-121N11.TJ RPCI-23 Mus musculus genomic clone
            RPCI-23-121N11, genomic survey sequence.
ACCESSION   AZ278916
VERSION     AZ278916.1  GI:9497818
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S.,
            Akinret,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de
            Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 121  row: N  column: 11
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-121N11"
                     /sex="Female"
                     /lab_host="DH10B"
                     /clone_lib="RPCI-23"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).   "
ORIGIN

Alignment Scores:
Pred. No.:              1.63e+03        Length:         291
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     9               Gaps:           0

US-09-858-332G-28 (1-6) x AZ278916 (1-291)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         32 CACAATCATAATCATAAT 15

RESULT 68
CL960835
LOCUS       CL960835                291 bp    DNA     linear   GSS 21-SEP-2004
DEFINITION  OsIFCC037313 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL960835
VERSION     CL960835.1  GI:52376381
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 291)
  AUTHORS   Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
            Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
            Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
```

```
COMMENT     Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1. .291
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

Alignment Scores:
Pred. No.:              1.63e+03        Length:         291
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     10              Gaps:           0

US-09-858-332G-28 (1-6) x CL960835 (1-291)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        157 CACAACCACAACCACAAC 174

RESULT 69
AJ645412
LOCUS       AJ645412                292 bp    mRNA    linear   EST 28-JAN-2005
DEFINITION  AJ645412 Populus tremula x P. tremuloides/Amanita muscaria mixed
            EST library Populus tremula x P. tremuloides/Amanita muscaria mixed
            EST library cDNA clone ptamabc210060a12, mRNA sequence.
ACCESSION   AJ645412
VERSION     AJ645412.1  GI:58314047
KEYWORDS    EST.
SOURCE      Populus tremula x P. tremuloides/Amanita muscaria mixed EST library
  ORGANISM  Populus tremula x P. tremuloides/Amanita muscaria mixed EST library
            Eukaryota; mixed EST libraries.
REFERENCE   1  (bases 1 to 292)
  AUTHORS   Bekel,T., Meyer,F., Kuester,H., Manthey,K., Perlick,A., Puehler,A.,
            Bock,A. and Nehls,U.
  TITLE     ESTs of fully developed ectomycorrhizas formed between Populus
            tremula x tremuloides and Amanita muscaria
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Bekel T.
            Physiologische Oekologie der Pflanzen
            Eberhard-Karls-Universitaet
            Auf der Morgenstelle 1, Tuebingen, BW 72076, Germany.
FEATURES             Location/Qualifiers
     source          1. .292
                     /organism="Populus tremula x P. tremuloides/Amanita
                     muscaria mixed EST library"
                     /mol_type="mRNA"
                     /db_xref="taxon:143775"
                     /clone="ptamabc210060a12"
                     /clone_lib="Populus tremula x P. tremuloides/Amanita
                     muscaria mixed EST library"
ORIGIN

Alignment Scores:
Pred. No.:              1.63e+03        Length:         292
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AJ645412 (1-292)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         71 CACAACCACAACCACAAT 88

RESULT 70
CO301685/c
LOCUS       CO301685                295 bp    mRNA    linear   EST 25-JUN-2004
DEFINITION  EK181910.5prime Exelixis FlyTag CK01 pCDNA-SK+ Drosophila
            melanogaster cDNA clone EK181910 5, mRNA sequence.
ACCESSION   CO301685
VERSION     CO301685.1  GI:49223381
KEYWORDS    EST.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 295)
  AUTHORS   Kopczynski,C., Platt,D., Campbell,J., Muzong,C., Laufer,A.,
            Peterson,E. and Swimmer,C.
  TITLE     Exelixis FlyTag EST Project CK01 Library
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Based upon one or more reads of this clone where vector sequence
            was present at both ends, this clone has been determined to contain
            contain a cDNA insert on the order of 600-1000 bases.
            Plate: EK.1819  row: A  column: 10
            High quality sequence stop: 293.
FEATURES             Location/Qualifiers
     source          1. .295
                     /organism="Drosophila melanogaster"
                     /mol_type="mRNA"
                     /db_xref="taxon:7227"
                     /clone="EK181910"
                     /clone_lib="Exelixis FlyTag CK01 pCDNA-SK+"
                     /note="Organ: mixed stage embryos, imaginal disks, and
                     adult heads; Vector: pCDNA-SK+; Site_1: NotI; Site_2:
                     XhoI; Random primed, normalized library from mixed stage
                     embryos, imaginal disks, and adult heads."
ORIGIN

Alignment Scores:
Pred. No.:              1.65e+03        Length:         295
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     7               Gaps:           0

US-09-858-332G-28 (1-6) x CO301685 (1-295)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        169 CATAATCATAATCATAAT 152

RESULT 71
BW876703/c
LOCUS       BW876703                297 bp    mRNA    linear   EST 24-MAY-2005
DEFINITION  BW876703 Amphioxus Branchiostoma floridae unpublished cDNA library,
            neurula whole animal Branchiostoma floridae cDNA clone bbne120h22
            5', mRNA sequence.
ACCESSION   BW876703
VERSION     BW876703.1  GI:66491380
KEYWORDS    EST.
SOURCE      Branchiostoma floridae (Florida lancelet)
  ORGANISM  Branchiostoma floridae
```

```
            Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
            Branchiostoma.
REFERENCE   1  (bases 1 to 297)
  AUTHORS   Yu,J., Holland,L.Z., Shin-i,T., Kohara,Y., Satou,Y. and Satoh,N.
  TITLE     Expressed genes in Branchiostoma floridae
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .297
                     /organism="Branchiostoma floridae"
                     /mol_type="mRNA"
                     /db_xref="taxon:7739"
                     /clone="bbne120h22"
                     /tissue_type="whole animal"
                     /dev_stage="neurula"
                     /clone_lib="Amphioxus Branchiostoma floridae unpublished
                     cDNA library, neurula whole animal"
ORIGIN

Alignment Scores:
Pred. No.:              1.66e+03        Length:         297
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     5               Gaps:           0

US-09-858-332G-28 (1-6) x BW876703 (1-297)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        175 CATAATCATAATCATAAT 158

RESULT 72
BF547803
LOCUS       BF547803                 298 bp    mRNA    linear   EST 11-DEC-2000
DEFINITION  UI-R-E0-bw-f-10-0-UI.r1 UI-R-E0 Rattus norvegicus cDNA clone
            UI-R-E0-bw-f-10-0-UI 5', mRNA sequence.
ACCESSION   BF547803
VERSION     BF547803.1  GI:11638910
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 298)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
   PUBMED   8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road , 4156  MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            clones will be available through Research Genetics (www.resgen.com)
            This clone is also available through the I.M.A.G.E. Consortium at
            LLNL (info@image.llnl.gov). IMAGE ID= 1769930 The following
            repetitive elements were found in this cDNA sequence: 227-279,
            >(CAA)n#Simple_repeat
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .298
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-E0-bw-f-10-0-UI"
                     /dev_stage="embryonic"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="UI-R-E0"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: NotI; Site_2: EcoRI; This library
                     consists of a mixture of indivIdually tagged normalized
                     libraries constructed from 8, 12 and 18-day embryo.  The
                     tag is a string  of 3-5 nucleotides present between the
                     Not I site and the oligo-dT track  which allows
                     identification of the library of origin of a clone within
                     the mixture."
ORIGIN

Alignment Scores:
Pred. No.:              1.67e+03        Length:         298
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-858-332G-28 (1-6) x BF547803 (1-298)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        239 CACAACCATAACCACAAC 256

RESULT 73
CD663439
LOCUS       CD663439                 298 bp    mRNA    linear   EST 23-JUN-2003
DEFINITION  UCRHV18_05ca01_b1 Drought-stressed Dicktoo barley epidermis cDNA
            library Hordeum vulgare subsp. vulgare cDNA clone UCRHV18_05ca01,
            mRNA sequence.
ACCESSION   CD663439
VERSION     CD663439.1  GI:32154703
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 298)
  AUTHORS   Wahid,A., Close,T.J., Fenton,R.D., Wanamaker,S., Collura,K.,
            Feuerbacher,O., Kim,H.R., Kudrna,D., Wing,R. and Yu,Y.
  TITLE     Drought-stressed barley leaf epidermis cDNA sequences
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Timothy J. Close
            Department of Botany & Plant Sciences
            University of California
            Riverside, CA 92521-0124, USA
            Tel: 909-787-3318
            Fax: 909-787-4437
            Email: timothy.close@ucr.edu
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .298
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Dicktoo"
                     /sub_species="vulgare"
                     /db_xref="taxon:112509"
                     /clone="UCRHV18_05ca01"
                     /tissue_type="lower leaf epidermis"
                     /dev_stage="1-2 week seedlings"
                     /lab_host="E. coli TJC121"
                     /clone_lib="Drought-stressed Dicktoo barley epidermis cDNA
                     library"
```

/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site_1: EcoRI; Site_2: XhoI; Seeds of barley (Hordeum vulgare L. cv. Dicktoo) were germinated in dishpans containing UCR-mix soil. Seedlings were kept in a growth chamber at 20C (day/night) and allowed to grow at 60-70% soil moisture content. After 3-4 days, the water was withheld in order to apply drought until the soil moisture content was reduced to 10-12%, which took another 3-4 days. At the time of extraction of epidermis, leaf water and osmotic potentials and extracted lower epidermis osmotic potentials had dropped to 15.9%, 20.2% and 24.7% of controls, respectively, as determined using a vapor pressure osmometer (Model 5100C, Wescor, Inc., Logan, UT). Epidermal strips were quickly peeled off of seedlings and immediately submerged in liquid nitrogen. About 15-20 g of epidermal tissue was collected and used to extract total RNA. Total RNA was extracted using Concert Plant RNA Reagent (Invitrogen 12322-012). Poly(A) RNA was purified using PolyATrack mRNA Isolation System IV (Promega). A cDNA library was made using a Uni-ZAP cDNA synthesis kit (Stratagene). A total of 1 million primary lambda cDNA clones were mass-excised in vivo to give a population of pBluescript SK(-) phagemids. All steps to this point were performed in the TJ Close lab at the University of California, Riverside by A. Wahid with some assistance from R.D. Fenton. Phagemids were plated on the TJC121 host strain, plasmid DNA purified, cDNA clones archived, and DNA sequences determined using the T7 primer (mainly 3' end reads) using an ABI3730 at the Arizona Genomics Institute, University of Arizona (Collura, Feuerbacher, Kim, Kudrna, Wing, Yu). Chromatogram files were transmitted to UC Riverside (by Yu), then processed at UC Riverside (by Wanamaker) using the HarvEST pipeline (http://harvest.ucr.edu) to remove vector and cloning oligo sequences and various contaminants, and to trim to a high quality region. Sequences that retained a phred 17 region of at least 100 bases were deposited to GenBank."

ORIGIN

Alignment Scores:

Pred. No.:	1.67e+03	Length:	298
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-09-858-332G-28 (1-6) x CD663439 (1-298)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        184 CATAACCATAATCATAAC 201
```

RESULT 74
T10142
LOCUS T10142 299 bp mRNA linear EST 13-APR-1994
DEFINITION seq1163 b4HB3MA Cot8-HAP-Ft Homo sapiens cDNA clone b4HB3MA-COT8-HAP-Ft152 3', mRNA sequence.
ACCESSION T10142
VERSION T10142.1 GI:471491
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 299)
AUTHORS Soares,M.B., Bonaldo,M.F., Jelenc,P., Su,L., Lawton,L and Efstratiadis,A.
TITLE Construction and characterization of a normalized cDNA library
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 91, 9228-9232 (1994)
PUBMED 7937745
COMMENT Contact: Bento Soares
Columbia University
Department of Psychiatry,722 W 168th Street, Unit #41, New York,NY 10032
Tel: 2129602313
Fax: 2127813577
Email: cuc@cuccfa.ccc.columbia.edu
antisense, Ampicillin
Seq primer: M13 Forward (Universal).
FEATURES Location/Qualifiers
source 1. .299
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="b4HB3MA-COT8-HAP-Ft152"
/lab_host="E. coli"
/clone_lib="b4HB3MA Cot8-HAP-Ft"
/note="Vector: Lafmid BA; Site_1: Not I; Site_2: Hind III; Size-selected cDNA from polyA+ RNA from human brain. 3-month old neonate.This library is the result of an attempted normalization of library b4HB3MA. "
ORIGIN

Alignment Scores:

Pred. No.:	1.67e+03	Length:	299
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	8	Gaps:	0

US-09-858-332G-28 (1-6) x T10142 (1-299)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        194 CATAATCATAATCATAAT 211
```

RESULT 75
BZ166889/c
LOCUS BZ166889 299 bp DNA linear GSS 11-OCT-2002
DEFINITION CH230-377D23.TJ CHORI-230 Segment 2 Rattus norvegicus genomic clone CH230-377D23, genomic survey sequence.
ACCESSION BZ166889
VERSION BZ166889.1 GI:23807953
KEYWORDS GSS.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 299)
AUTHORS Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de Jong,P. and Fraser,C.M.
TITLE Rat BAC End Sequences from Library CHORI-230 MboI segment
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: CH230-377D23.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230 (http://www.chori.org/bacpac/rat230.htm). For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.chori.org/bacpac/or ering_information.htm). BAC end page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 377 row: D column: 23
Seq primer: SP6
Class: BAC ends.
FEATURES Location/Qualifiers

```
     source          1. .299
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-377D23"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 2"
                     /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
ORIGIN

Alignment Scores:
Pred. No.:              1.67e+03        Length:         299
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     9               Gaps:           0

US-09-858-332G-28 (1-6) x BZ166889 (1-299)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        297 CACAACCACAACCACAAC 280

RESULT 76
AJ592259
LOCUS       AJ592259                299 bp    DNA     linear   GSS 15-JAN-2004
DEFINITION  Arabidopsis thaliana T-DNA flanking sequence, right border, clone
            610D11, genomic survey sequence.
ACCESSION   AJ592259
VERSION     AJ592259.1  GI:37941883
KEYWORDS    GSS; right border; T-DNA flanking sequence.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Brunaud,V., Balzergue,S., Dubreucq,B., Aubourg,S., Samson,F.,
            Chauvin,S., Bechtold,N., Cruaud,C., DeRose,R., Pelletier,G.,
            Lepiniec,L., Caboche,M. and Lecharny,A.
  TITLE     T-DNA integration into the Arabidopsis genome depends on sequences
            of pre-insertion sites
  JOURNAL   EMBO Rep. 3 (12), 1152-1157 (2002)
   PUBMED   12446565
REFERENCE   2  (bases 1 to 299)
  AUTHORS   Balzergue,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2003) Balzergue S., UMRGV, INRA/CNRS, 2 rue
            Gaston Cremieux, 91057 Evry cedex, FRANCE
COMMENT     PCR was performed on DNA from transformants of Arabidopsis thaliana
            plants from INRA (Versailles). The DNA fragment(s) resulting from
            the PCR were directly sequenced from the left or the right border
            to determine the genomic sequence flanking the insertion. T-DNA
            derived sequences were removed. Information to order the
            corresponding mutant line and a link to a database providing a
            graphical display of the insertion site are available at
            http://dbsgap.versailles.inra.fr/publiclines/. This sequence has
            been generated in the framework of the French plant genomics
            program 'Genoplante' (http://www.genoplante.com and
            http://genoplante-info.infobiogen.fr).
FEATURES                Location/Qualifiers
     source             1. .299
                        /organism="Arabidopsis thaliana"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:3702"
                        /clone="610D11"
                        /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
                        /ecotype="Wassilewskija"
     misc_feature       1. .299
                        /note="T-DNA flanking sequence
                        right border"
ORIGIN

Alignment Scores:
Pred. No.:              1.67e+03        Length:         299
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     10              Gaps:           0

US-09-858-332G-28 (1-6) x AJ592259 (1-299)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        116 CATAATCATAACCATAAC 133

RESULT 77
AJ473189/c
LOCUS       AJ473189                300 bp    mRNA    linear   EST 24-MAY-2002
DEFINITION  AJ473189 S00008 Hordeum vulgare cDNA clone S0000800175E01F1, mRNA
            sequence.
ACCESSION   AJ473189
VERSION     AJ473189.1  GI:21189145
KEYWORDS    EST.
SOURCE      Hordeum vulgare
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 300)
  AUTHORS   Saren,A.-M., Tanskanen,J., Paulin,L. and Schulman,A.H.
  TITLE     Barley EST's
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Schulman AH
            Institute of Biotechnology
            University of Helsinki
            P.O.Box 56 (Viikinkaari 6A), University of Helsinki FIN-00014,
            Finland.
FEATURES                Location/Qualifiers
     source             1. .300
                        /organism="Hordeum vulgare"
                        /mol_type="mRNA"
                        /db_xref="taxon:4513"
                        /clone="S0000800175E01F1"
                        /tissue_type="Callus"
                        /clone_lib="S00008"
                        /note="Callus K19"
ORIGIN

Alignment Scores:
Pred. No.:              1.68e+03        Length:         300
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AJ473189 (1-300)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        139 CATAACCATAATCATAAC 122

RESULT 78
AI485260
LOCUS       AI485260                305 bp    mRNA    linear   EST 18-MAY-2001
DEFINITION  EST243564 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
            cLED4L22, mRNA sequence.
ACCESSION   AI485260
VERSION     AI485260.1  GI:4380631
```

KEYWORDS EST.
SOURCE Lycopersicon esculentum (Solanum lycopersicum)
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 305)
AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
TITLE Generation of ESTs from tomato carpel tissue
JOURNAL Unpublished (1999)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html.
FEATURES Location/Qualifiers
source 1. .305
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLED4L22"
/tissue_type="carpel"
/dev_stage="5 days pre-anthesis to 5 days post-anthesis"
/lab_host="XL1-Blue MRF'"
/clone_lib="tomato ovary, TAMU"
/note="Vector: pBlueScript SK(-); Site_1: EcoR1; Site_2: Xho1; cLED - Tomato Carpel EST Library. OligodT-primed and directionally cloned cDNA in vector Lamda ZAP II with 5' and 3' ends located at the EcoRI and XhoI sites, respectively."
ORIGIN

Alignment Scores:

Pred. No.:	1.71e+03	Length:	305
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0

US-09-858-332G-28 (1-6) x AI485260 (1-305)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         31 CACAACCACAATCACAAC 48
```

RESULT 79
BE956148/c
LOCUS BE956148 306 bp mRNA linear EST 04-OCT-2000
DEFINITION UI-M-BH4-bba-e-11-0-UI.s1 NIH_BMAP_M_S5 Mus musculus cDNA clone UI-M-BH4-bba-e-11-0-UI 3', mRNA sequence.
ACCESSION BE956148
VERSION BE956148.1 GI:10600402
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 306)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
PUBMED 8889548
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
Oligo-dT track not found, Not I site shown in beginning of sequence is likely internal to the message. cDNA Library Preparation: M.B. Soares Lab Clone distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements
Seq primer: M13 Forward
POLYA=No.
FEATURES Location/Qualifiers
source 1. .306
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BH4-bba-e-11-0-UI"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NIH_BMAP_M_S5"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The NIH_BMAP_M_S5 library is a subtracted library of a series, ultimately derived from a mixture of individually tagged normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hipoccampus) after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. For a detailed description of the library from which this clone was derived, please visit our web site at brainest.eng.uiowa.edu.
TAG_SEQ=None found"
ORIGIN

Alignment Scores:

Pred. No.:	1.72e+03	Length:	306
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-28 (1-6) x BE956148 (1-306)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        301 CACAACCACAACCACAAC 284
```

RESULT 80
BG046168
LOCUS BG046168 307 bp mRNA linear EST 23-JUL-2004
DEFINITION saa49e08.y1 Gm-c1060 Glycine soja cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1060-207 5', mRNA sequence.
ACCESSION BG046168
VERSION BG046168.1 GI:12494649
KEYWORDS EST.
SOURCE Glycine soja
ORGANISM Glycine soja
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE 1 (bases 1 to 307)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project

JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
When it has been determined, an EST from the other end of this clone is listed in the 'Other ESTs on clone' field. This clone is available through: Biogenetic Services, 801 32nd Ave. Brookings, SD 57006 USA (phone: 800 423 4163; email: info@biogeneticservices.com)
High quality sequence stop: 242.
FEATURES Location/Qualifiers
source 1. .307
/organism="Glycine soja"
/mol_type="mRNA"
/db_xref="taxon:3848"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1060-207"
/tissue_type="Root, 2 week seedlings"
/lab_host="DH10B"
/clone_lib="Gm-c1060"
/note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2: XhoI; The cDNA library was constructed from mRNA isolated from root tissue of 2 week old seedlings for PI468916. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (Gibco BRL). This library was constructed in the laboratory of Dr. Randy Shoemaker at Iowa state university."
ORIGIN

Alignment Scores:

Pred. No.:	1.72e+03	Length:	307
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-28 (1-6) x BG046168 (1-307)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         50 CATAACCATAACCATAAC 67
```

RESULT 81
BY356096/c
LOCUS BY356096 308 bp mRNA linear EST 12-DEC-2002
DEFINITION BY356096 RIKEN full-length enriched, adult male tongue Mus musculus cDNA clone 2310042N04 3', mRNA sequence.
ACCESSION BY356096
VERSION BY356096.1 GI:26585584
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 308)
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
PUBMED 12466851
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1. .308
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="2310042N04"
/sex="male"
/tissue_type="tongue"
/dev_stage="adult"
/lab_host="SOLR"
/clone_lib="RIKEN full-length enriched, adult male tongue"
/note="Site_1: XhoI; Site_2: SstI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was

prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. "

ORIGIN

Alignment Scores:

Pred. No.:	1.73e+03	Length:	308
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	5	Gaps:	0

US-09-858-332G-28 (1-6) x BY356096 (1-308)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         72 CACAACCATAACCACAAC 55
```

RESULT 82
AQ250124/c
LOCUS AQ250124 310 bp DNA linear GSS 06-OCT-1998
DEFINITION T21D6-Sp6.2 TAMU Arabidopsis thaliana genomic clone T21D6, genomic survey sequence.
ACCESSION AQ250124
VERSION AQ250124.1 GI:3700207
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 310)
AUTHORS Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and Ecker,J.
TITLE BAC End Sequences at ATGC
JOURNAL Unpublished (1997)
COMMENT Contact: Ecker J.
Arabidopsis Thaliana Genome Center
University of Pennsylvania
Dept. of Biology, University of Pennsylvania, Philadelphia, PA 19104
Tel: 215-898-9384
Fax: 215-898-8780
Email: jecker@atgenome.bio.upenn.edu
Seq primer: Sp6
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .310
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Columbia"
/db_xref="taxon:3702"
/clone="T21D6"
/sex="hermaphrodite"
/clone_lib="TAMU"
/note="Vector: BeloBACII; Site_1: HindIII; Site_2: HindIII; Produced by Rod Wing"
ORIGIN

Alignment Scores:

Pred. No.:	1.74e+03	Length:	310
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

US-09-858-332G-28 (1-6) x AQ250124 (1-310)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        170 CACAACCACAACCATAAC 153
```

RESULT 83
BF486624
LOCUS BF486624 311 bp mRNA linear EST 19-APR-2001
DEFINITION AT21076.5prime AT Drosophila melanogaster adult testes pOTB7 Drosophila melanogaster cDNA clone AT21076 5 similar to CG14732: FBan0014732 located on: 3R 87A6-87A6;: 04/09/2001, mRNA sequence.
ACCESSION BF486624
VERSION BF486624.2 GI:13698497
KEYWORDS EST.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 311)
AUTHORS Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Baxter,E., Berman,B., Carlson,J., Champe,M., Chavez,C., Chew,M., Dorsett,V., Farfan,D., Frise,E., George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Miranda,A., Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S., Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin,G.M.
TITLE BDGP/HHMI AT Drosophila EST Project
JOURNAL Unpublished (2000)
COMMENT On Dec 6, 2000 this sequence version replaced gi:11569925.
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
hit genomic AE003693: arm:3R [7544898,7771380]
estimated-cyto:86F7-87A8: 04/09/2001
Plate: AT.210 row: G column: 4
High quality sequence stop: 303.
FEATURES Location/Qualifiers
source 1. .311
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="AT21076"
/sex="male"
/dev_stage="0-3 day old Ore-R males"
/lab_host="Plates AT.10-AT.120: DH5-alpha. Plates AT.121-AT.319: DH5-alpha TonA"
/clone_lib="AT Drosophila melanogaster adult testes pOTB7"
/note="Organ: ADULT testes; Vector: pOTB7; Site_1: EcoRI; Site_2: Xho1; The mRNA for the testis library was made from testes and seminal vessicles hand dissected from 0-3 day old Ore-R males. RNA kindly provided by the lab of Margaret Fuller. Sized fractionated cDNAs were directly ligated into pOTB7. Plasmid cDNA library."
ORIGIN

Alignment Scores:

Pred. No.:	1.75e+03	Length:	311
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-28 (1-6) x BF486624 (1-311)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         96 CACAACCATAATCACAAT 113
```

RESULT 84
AV560899/c

```
LOCUS       AV560899                313 bp    mRNA    linear   EST 23-FEB-2004
DEFINITION  AV560899 Arabidopsis thaliana green siliques Columbia Arabidopsis
            thaliana cDNA clone SQ142c11F 3', mRNA sequence.
ACCESSION   AV560899
VERSION     AV560899.1  GI:8732325
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 313)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     A large scale analysis of cDNA in Arabidopsis thaliana: Generation
            of 12,028 non-redundant expressed sequence tags from normalized and
            size-selected cDNA libraries
  JOURNAL   DNA Res. 7 (3), 175-180 (2000)
   PUBMED   10907847
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .313
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /ecotype="Columbia"
                     /db_xref="taxon:3702"
                     /clone="SQ142c11F"
                     /tissue_type="green siliques"
                     /clone_lib="Arabidopsis thaliana green siliques Columbia"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
ORIGIN

Alignment Scores:
Pred. No.:              1.76e+03        Length:         313
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AV560899 (1-313)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        291 CACAACCACAACCACAAC 274

RESULT 85
D43098/c
LOCUS       D43098                  313 bp    mRNA    linear   EST 07-DEC-2004
DEFINITION  D43098 Rice callus cDNA (H.Uchimiya) Oryza sativa cDNA clone AD623,
            mRNA sequence.
ACCESSION   D43098
VERSION     D43098.1  GI:3107358
KEYWORDS    EST.
SOURCE      Oryza sativa
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 313)
  AUTHORS   Uchimiya,H.
  TITLE     On nucleotide sequence of Oryza sativa
  JOURNAL   Unpublished (1994)
COMMENT     Contact: Hirofumi Uchimiya
            Institute of Mol. & Cell. Bioscience, Department of Cellular
            Function
            The University of Tokyo
            1-1-1 Yayoi, Bunkyo-ku, Tokyo 113, Japan
            Tel: 81-3-5841-7844(ex.7844)
            Fax: 81-3-5841-8466
            Email: uchimiya@iam.u-tokyo.ac.jp
            PROJECT= Uchimiya
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .313
                     /organism="Oryza sativa"
                     /mol_type="mRNA"
                     /db_xref="taxon:4530"
                     /clone="AD623"
                     /tissue_type="callus"
                     /clone_lib="Rice callus cDNA (H.Uchimiya)"
ORIGIN

Alignment Scores:
Pred. No.:              1.76e+03        Length:         313
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     8               Gaps:           0

US-09-858-332G-28 (1-6) x D43098 (1-313)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        193 CACAACCACAACCACAAC 176

RESULT 86
BJ351548/c
LOCUS       BJ351548                314 bp    mRNA    linear   EST 06-MAR-2002
DEFINITION  BJ351548 Dictyostelium discoideum cDNA library, AF Dictyostelium
            discoideum cDNA clone dda43h24 3', mRNA sequence.
ACCESSION   BJ351548
VERSION     BJ351548.1  GI:19222094
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 314)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the aggregation
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .314
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dda43h24"
                     /sex="mat A"
                     /dev_stage="Aggregation stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN

Alignment Scores:
Pred. No.:              1.76e+03        Length:         314
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-858-332G-28 (1-6) x BJ351548 (1-314)
```

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        269 CACAACCACAACCACAAC 252

RESULT 87
AI483545
LOCUS       AI483545                315 bp    mRNA    linear   EST 18-MAY-2001
DEFINITION  EST249394 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
            cLED25M3, mRNA sequence.
ACCESSION   AI483545
VERSION     AI483545.1  GI:4387469
KEYWORDS    EST.
SOURCE      Lycopersicon esculentum (Solanum lycopersicum)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 315)
  AUTHORS   Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T.,
            Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B.,
            Fujii,C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C.,
            Martin,G.B., Tanksley,S.D. and Giovannoni,J.
  TITLE     Generation of ESTs from tomato carpel tissue
  JOURNAL   Unpublished (1999)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html.
FEATURES             Location/Qualifiers
     source          1. .315
                     /organism="Lycopersicon esculentum"
                     /mol_type="mRNA"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cLED25M3"
                     /tissue_type="carpel"
                     /dev_stage="5 days pre-anthesis to 5 days post-anthesis"
                     /lab_host="XL1-Blue MRF'"
                     /clone_lib="tomato ovary, TAMU"
                     /note="Vector: pBlueScript SK(-); Site_1: EcoR1; Site_2:
                     Xho1; cLED - Tomato Carpel EST Library. OligodT-primed and
                     directionally cloned  cDNA in vector Lamda ZAP II with 5'
                     and 3' ends located at the EcoRI and  XhoI sites,
                     respectively."
ORIGIN

Alignment Scores:
Pred. No.:                1.77e+03        Length:         315
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       1               Gaps:           0

US-09-858-332G-28 (1-6) x AI483545 (1-315)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         33 CACAACCACAATCACAAC 50

RESULT 88
CK116691
LOCUS       CK116691                315 bp    mRNA    linear   EST 01-DEC-2003
DEFINITION  B015P72 Hybrid aspen plasmid library Populus tremula x Populus
            tremuloides cDNA clone B015P72 5', mRNA sequence.
ACCESSION   CK116691
VERSION     CK116691.1  GI:38601016
KEYWORDS    EST.
SOURCE      Populus tremula x Populus tremuloides
  ORGANISM  Populus tremula x Populus tremuloides
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Malpighiales; Salicaceae; Saliceae; Populus.
REFERENCE   1  (bases 1 to 315)
  AUTHORS   Sterky,F., Bhalerao,R.R., Unneberg,P., Segerman,B., Nilsson,P.,
            Brunner,A.M., Charbonnel-Campaa,L., Lindvall,J.J., Tandre,K.,
            Strauss,S.H., Sundberg,B., Gustafsson,P., Uhlen,M., Bhalerao,R.P.,
            Nilsson,O., Sandberg,G., Karlsson,J., Lundeberg,J. and Jansson,S.
  TITLE     A Populus EST resource for plant functional genomics
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 101 (38), 13951-13956 (2004)
  PUBMED    15353603
COMMENT     Contact: Bo Segerman
            Umea Plant Science Center, Department of Plant Physiology
            Umea University
            901 87 Umea, Sweden
            Tel: +46 90 786 5279
            Fax: +46 90 786 6676
            Email: bo.segerman@plantphys.umu.se.
FEATURES             Location/Qualifiers
     source          1. .315
                     /organism="Populus tremula x Populus tremuloides"
                     /mol_type="mRNA"
                     /db_xref="taxon:47664"
                     /clone="B015P72"
                     /tissue_type="Cambial region"
                     /dev_stage="1.5 m actively growing tree"
                     /lab_host="E.coli"
                     /clone_lib="Hybrid aspen plasmid library"
                     /note="Vector: pBluescript SK; Site_1: SalI; Site_2: NotI;
                     Cambial region tissues, including developing xylem, the
                     meristematic  cambial zone and the developing and mature
                     phloem, was harvested  from 1.5 m actively growing trees.
                     cDNA was prepared and cloned  into lambda gt22a. DNA was
                     isolated and subcloned into pBluescript SK using SalI and
                     NotI restriction enzymes."
ORIGIN

Alignment Scores:
Pred. No.:                1.77e+03        Length:         315
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       7               Gaps:           0

US-09-858-332G-28 (1-6) x CK116691 (1-315)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        173 CACAACCACAACCACAAC 190

RESULT 89
H35592
LOCUS       H35592                  316 bp    mRNA    linear   EST 02-APR-1998
DEFINITION  EST110865 Rat PC-12 cells, NGF-treated (9 days) Rattus norvegicus
            cDNA clone RPNBN84 similar to H.sapiens hypothetical protein, 13
            kDa, mRNA sequence.
ACCESSION   H35592
VERSION     H35592.1  GI:981009
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 316)
  AUTHORS   Lee,N.H., Weinstock,K.G., Kirkness,E.F., Earle-Hughes,J.A.,
            Fuldner,R.A., Marmaras,S., Glodek,A., Gocayne,J.D., Adams,M.D.,
            Kerlavage,A.R., Fraser,C.M. and Venter,J.C.
  TITLE     Comparative expressed-sequence-tag analysis of differential gene
            expression profiles in PC-12 cells before and after nerve growth
            factor treatment
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 92, 8303-8307 (1995)
  PUBMED    7667285
COMMENT     Other_ESTs: TC725
```

```
            Contact: Lee, NH
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            For clone availability please contact  the TIGR Database
            (tdbinfo@tdb.tigr.org) TC (Tentative Consensus) numbers represent
            assemblies of ESTs.
FEATURES              Location/Qualifiers
     source           1. .316
                      /organism="Rattus norvegicus"
                      /mol_type="mRNA"
                      /db_xref="ATCC (inhost):2005025"
                      /db_xref="taxon:10116"
                      /clone="RPNBN84"
                      /clone_lib="Rat PC-12 cells, NGF-treated (9 days)"
                      /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                      XhoI; poly(A)+ RNA was purified from 9-day NGF treated
                      PC12 cells. cDNA was constructed using an oligo-dT primer
                      and directionally cloned using the Lambda ZAP II Vector
                      Kit by Stratagene"
ORIGIN

Alignment Scores:
Pred. No.:              1.78e+03       Length:         316
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     8              Gaps:           0

US-09-858-332G-28 (1-6) x H35592 (1-316)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        238 CACAACCATAACCACAAC 255

RESULT 90
AU227400
LOCUS       AU227400                317 bp    mRNA    linear   EST 23-APR-2002
DEFINITION  AU227400 RAFL15 Arabidopsis thaliana cDNA clone RAFL15-04-C12 3',
            mRNA sequence.
ACCESSION   AU227400
VERSION     AU227400.1  GI:19742047
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 317)
  AUTHORS   Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
            Akiyama,K., Enju,A., Oono,Y., Sakurai,T., Carninci,P., Kawai,J.,
            Itoh,M., Ishii,Y., Arakawa,T., Shibata,K., Shinagawa,A.,
            Muramatsu,M., Hayashizaki,Y. and Shinozaki,K.
  TITLE     Large scale analysis of Arabidopsis full-length cDNA
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            An Arabidopsis full-length cDNA library was constructed essentially
            as reported previously (Seki et al., 1998).cDNA cleaved with BamHI
            and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et
            al., submitted for publication) digested with BamHI and SalI. This
            clone is in a modified pBluescript vector. Please visit our web
            site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further
            details.
FEATURES              Location/Qualifiers
     source           1. .317
                      /organism="Arabidopsis thaliana"
                      /mol_type="mRNA"
                      /db_xref="taxon:3702"
                      /clone="RAFL15-04-C12"
                      /tissue_type="mixture of silique and flower"
                      /lab_host="DH10B"
                      /clone_lib="RAFL15"
                      /note="Site_1: BamHI; Site_2: SalI"
ORIGIN

Alignment Scores:
Pred. No.:              1.78e+03       Length:         317
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     1              Gaps:           0

US-09-858-332G-28 (1-6) x AU227400 (1-317)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         26 CATAATCATAATCATAAT 43

RESULT 91
D47535
LOCUS       D47535                  317 bp    mRNA    linear   EST 03-APR-2002
DEFINITION  RICS13103A Rice green shoot Oryza sativa (japonica cultivar-group)
            cDNA, mRNA sequence.
ACCESSION   D47535
VERSION     D47535.1  GI:701244
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 317)
  AUTHORS   Sasaki,T., Miyao,A. and Yamamoto,K.
  TITLE     Rice cDNA from callus 1995
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
            305-8602, Japan
            Tel: 81-298-38-7441
            Fax: 81-298-38-7468
            Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
            PROJECT ='RGP'.
FEATURES              Location/Qualifiers
     source           1. .317
                      /organism="Oryza sativa (japonica cultivar-group)"
                      /mol_type="mRNA"
                      /cultivar="Nipponbare"
                      /db_xref="taxon:39947"
                      /clone_lib="Rice green shoot"
                      /note="Green shoot (8 days old)"
ORIGIN

Alignment Scores:
Pred. No.:              1.78e+03       Length:         317
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     8              Gaps:           0

US-09-858-332G-28 (1-6) x D47535 (1-317)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        151 CACAACCACAACCACAAC 168
```

```
RESULT 92
R39818
LOCUS       R39818                   317 bp    mRNA    linear   EST 05-MAY-1995
DEFINITION  yf51a04.s1 Soares infant brain 1NIB Homo sapiens cDNA clone
            IMAGE:25336 3', mRNA sequence.
ACCESSION   R39818
VERSION     R39818.1  GI:797434
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 317)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1539
            High quality sequence stops: 86 Source: IMAGE Consortium, LLNL This
            clone is available royalty-free through LLNL ; contact the IMAGE
            Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1539    Std Error: 0.00
            Seq primer: -21m13
            High quality sequence stop: 86.
FEATURES             Location/Qualifiers
     source          1. .317
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="GDB:397683"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:25336"
                     /sex="female"
                     /dev_stage="73 days post natal"
                     /lab_host="DH10B (ampicillin resistant)"
                     /clone_lib="Soares infant brain 1NIB"
                     /note="Organ: whole brain; Vector: Lafmid BA; Site_1: Not
                     I; Site_2: Hind III; 1st strand cDNA was primed with a Not
                     I - oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTTT 3']
                     double-stranded cDNA was ligated to Hind III adaptors
                     (Pharmacia), digested with Not I and directionally cloned
                     into the Not I and Hind III sites of the Lafmid BA vector.
                     Library went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
ORIGIN

Alignment Scores:
Pred. No.:                1.78e+03        Length:         317
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       8               Gaps:           0

US-09-858-332G-28 (1-6) x R39818 (1-317)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        200 CATAATCATAATCATAAT 217

RESULT 93
CE715092/c
LOCUS       CE715092                 318 bp    DNA     linear   GSS 29-SEP-2003
DEFINITION  tigr-gss-dog-17000369508467 Dog Library Canis familiaris genomic,
            genomic survey sequence.
ACCESSION   CE715092
VERSION     CE715092.1  GI:37034530
KEYWORDS    GSS.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
            Canis.
REFERENCE   1  (bases 1 to 318)
  AUTHORS   Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
            Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
            Venter,J.C.
  TITLE     The dog genome: survey sequencing and comparative analysis
  JOURNAL   Science 301 (5641), 1898-1903 (2003)
  PUBMED    14512627
COMMENT     Contact: Kirkness EF
            The Institute for Genomic Research
            Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
            Rockville, MD 20850, USA
            Tel: 301-838-0200
            Fax: 301-838-0208
            Email: ekirknes@tigr.org
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .318
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /strain="Standard Poodle"
                     /db_xref="taxon:9615"
                     /clone_lib="Dog Library"
                     /note="Site_1: BstXI; Libraries were prepared from
                     peripheral blood"
ORIGIN

Alignment Scores:
Pred. No.:                1.79e+03        Length:         318
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       10              Gaps:           0

US-09-858-332G-28 (1-6) x CE715092 (1-318)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        254 CATAATCATAATCATAAT 237

RESULT 94
AU061961
LOCUS       AU061961                 319 bp    mRNA    linear   EST 20-MAY-1999
DEFINITION  AU061961 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
            discoideum cDNA clone SLG781, mRNA sequence.
ACCESSION   AU061961
VERSION     AU061961.1  GI:4883065
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 319)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
```

```
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp
            PROJECT = Dictyostelium discoideum cDNA project in Japan.
FEATURES                Location/Qualifiers
     source             1. .319
                        /organism="Dictyostelium discoideum"
                        /mol_type="mRNA"
                        /strain="AX4"
                        /db_xref="taxon:44689"
                        /clone="SLG781"
                        /dev_stage="slug"
                        /clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
ORIGIN

Alignment Scores:
Pred. No.:              1.79e+03        Length:         319
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AU061961 (1-319)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         85 CACAACCACAACCACAAC 102

RESULT 95
CN927588
LOCUS       CN927588                320 bp    mRNA    linear   EST 07-JUN-2004
DEFINITION  000530AEPA003857HT (AEPA) Pinkie expanding leaf Malus x domestica
            cDNA clone AEPA003857, mRNA sequence.
ACCESSION   CN927588
VERSION     CN927588.1  GI:48400401
KEYWORDS    EST.
SOURCE      Malus x domestica
  ORGANISM  Malus x domestica
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Rosales; Rosaceae; Maloideae; Malus.
REFERENCE   1  (bases 1 to 320)
  AUTHORS   Beuning,L., Bowen,J., Crowhurst,R., Gleave,A., Janssen,B.,
            McArtney,S., Newcomb,R., Ross,G., Snowden,K., Walton,E. and Yauk,Y.
  TITLE     HortResearch Apple EST Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Gleave,A.
            Sequencing Facility
            The Horticulture and Food Research Institute of New Zealand Ltd
            120 Mt Albert Rd, Mt Albert, Auckland, New Zealand
            Tel: 00 64 09 815 4200
            Fax: 00 64 09 815 4201
            Email: est@hortresearch.co.nz.
FEATURES                Location/Qualifiers
     source             1. .320
                        /organism="Malus x domestica"
                        /mol_type="mRNA"
                        /db_xref="taxon:3750"
                        /clone="AEPA003857"
                        /tissue_type="Leaf"
                        /dev_stage="Expanding"
                        /clone_lib="(AEPA) Pinkie expanding leaf"
                        /note="Vector: pBK-CMV; Library sequenced by Genesis
                        Research & Development"
ORIGIN

Alignment Scores:
Pred. No.:              1.8e+03         Length:         320
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     7               Gaps:           0

US-09-858-332G-28 (1-6) x CN927588 (1-320)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        185 CACAACCACAACCACAAC 202

RESULT 96
BQ984974/c
LOCUS       BQ984974                321 bp    mRNA    linear   EST 21-AUG-2002
DEFINITION  QGE5J07.yg.ab1 QG_EFGHJ lettuce serriola Lactuca sativa cDNA clone
            QGE5J07, mRNA sequence.
ACCESSION   BQ984974
VERSION     BQ984974.1  GI:22402499
KEYWORDS    EST.
SOURCE      Lactuca sativa
  ORGANISM  Lactuca sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
            Cichorieae; Lactuca.
REFERENCE   1  (bases 1 to 321)
  AUTHORS   Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J.,
            Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y.,
            Lai,Z., Church,S., Jackson,L. and Bradford,K.
  TITLE     Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
            belongs to contig QG_CA_Contig7522, see http://cgpdb.ucdavis.edu/
            for details.
            Plate: QGE5  row: J  column: 07.
FEATURES                Location/Qualifiers
     source             1. .321
                        /organism="Lactuca sativa"
                        /mol_type="mRNA"
                        /cultivar="L.serriola"
                        /db_xref="taxon:4236"
                        /clone="QGE5J07"
                        /lab_host="E.coli"
                        /clone_lib="QG_EFGHJ lettuce serriola"
                        /note="Vector: pBRcDNASfiAB; The library was constructed
                        from 10 different sources of RNA from a single genotype.
                        Separate cDNAs were generated using primers that
                        incorporated unique 5' and 3' tags to distinguish each
                        source of RNA. cDNAs were then pooled, size-fractionated,
                        directionally cloned into a custom medium-copy vector and
                        transformations made with four size classes to minimize
                        size bias. Details of each source of RNA and library
                        construction can be obtained at http://cgpdb.ucdavis.edu/
                        TAG_TISSUE=chemical induction
                        TAG_LIB=QG_EFGHJ lettuce serriola
                        TAG_SEQ=TGTAGCCGGG"
ORIGIN

Alignment Scores:
Pred. No.:              1.81e+03        Length:         321
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     5               Gaps:           0

US-09-858-332G-28 (1-6) x BQ984974 (1-321)

Qy          1 HisAsnHisAsnHisAsn 6
```

```
                    ||||||||||||||||||
Db        133 CACAATCACAATCACAAT 116

RESULT 97
CF985192
LOCUS       CF985192                321 bp    mRNA    linear   EST 04-FEB-2005
DEFINITION  24466rsicee_8102.y1 Oryza sativa cv. LYP9 tillering whole plant
            cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF985192
VERSION     CF985192.1  GI:58586884
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 321)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 321
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .321
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="tillering"
                     /clone_lib="Oryza sativa cv. LYP9 tillering whole plant
                     cDNA library"
ORIGIN

Alignment Scores:
Pred. No.:              1.81e+03        Length:         321
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     7               Gaps:           0

US-09-858-332G-28 (1-6) x CF985192 (1-321)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        222 CACAACCACAACCACAAC 239

RESULT 98
BU275786
LOCUS       BU275786                322 bp    mRNA    linear   EST 03-JAN-2003
DEFINITION  Cr_Emb_07E01_TEXF1 Convoluta roscoffensis embryos from Eva Jiminez
            Symsagittifera roscofensis cDNA clone Cr_Emb_07E01 5', mRNA
            sequence.
ACCESSION   BU275786
VERSION     BU275786.1  GI:27476556
KEYWORDS    EST.
SOURCE      Symsagittifera roscofensis
  ORGANISM  Symsagittifera roscofensis
            Eukaryota; Metazoa; Platyhelminthes; Turbellaria; Acoelomorpha;
            Acoela; Sagittiferidae; Symsagittifera.
REFERENCE   1  (bases 1 to 322)
  AUTHORS   Blaxter,M.L., Jiminez,E., Rosie,A., Aboobaker,A. and Salo,E.
  TITLE     A survey of genes expressed in the acoel flatworm Convoluta
            roscoffensis
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Blaxter ML
            Institute of Cell, Animal and Population Biology
            University of Edinburgh
            Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
            3JT, UK.
            Tel: +44 131 650 6760
            Fax: +44 131 670 5450
            Email: mark.blaxter@ed.ac.uk
            The library was prepared by Eva Jimenez, University of Barcelona,
            Spain The sequencing was performed by Eva Jimenez, Aziz Aboobaker
            and Ailie Rosie Edinburgh Phylogenomics Programme, ICAPB,
            Edinburgh, UK.The sequence contained a PolyA tail (trimmed)
            PCR PRimers
            FORWARD: T7PL
            BACKWARD: triple 5 Exseq
            Plate: 07  row: E  column: 01
            Seq primer: TEXF1
            High quality sequence stop: 322.
FEATURES             Location/Qualifiers
     source          1. .322
                     /organism="Symsagittifera roscofensis"
                     /mol_type="mRNA"
                     /db_xref="taxon:84072"
                     /clone="Cr_Emb_07E01"
                     /sex="mixed"
                     /dev_stage="embryo"
                     /clone_lib="Convoluta roscoffensis embryos from Eva
                     Jiminez"
                     /note="Vector: lambda triplex; Site_1: EcoRI; Site_2:
                     HinDIII; The library was prepared by Eva Jimenez,
                     University of Barcelona, Spain The sequencing was
                     performed by Eva Jimenez,  Aziz Aboobaker and Ailie Rosie,
                     Edinburgh Phylogenomics Programme,  ICAPB, Edinburgh, UK."
ORIGIN

Alignment Scores:
Pred. No.:              1.81e+03        Length:         322
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     5               Gaps:           0

US-09-858-332G-28 (1-6) x BU275786 (1-322)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        265 CATAACCATAACCATAAC 282

RESULT 99
CG685934
LOCUS       CG685934                322 bp    DNA     linear   GSS 08-OCT-2003
DEFINITION  ZMMBBc0162P14f ZMMBBc (EcoRI) Zea mays genomic clone ZMMBBc0162P14
            5', genomic survey sequence.
ACCESSION   CG685934
VERSION     CG685934.1  GI:37583322
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 322)
  AUTHORS   Bharti,A.K., Young,S., Kavchok,S., Keizer,G., Bronzino,A.C.,
```

```
            Zohovetz,V., Fuks,G., Yu,Y., Wing,R. and Messing,J.
  TITLE     Sequencing of the maize genome at PGIR (2003c)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Bharti,A.K.
            Dr.Joachim Messing's lab
            The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers
            University
            190 Frelinghuysen Road, Piscataway, NJ 08854, USA
            Tel: 732 445 3801
            Fax: 732 445 5735
            Email: bharti@waksman.rutgers.edu
            Seq primer: T7
            Class: BAC ends
            High quality sequence start: 172.
FEATURES             Location/Qualifiers
     source          1. .322
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBBc0162P14"
                     /lab_host="E. coli DH10B"
                     /clone_lib="ZMMBBc (EcoRI)"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI"
ORIGIN

Alignment Scores:
Pred. No.:                1.81e+03        Length:         322
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       10              Gaps:           0

US-09-858-332G-28 (1-6) x CG685934 (1-322)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        289 CATAATCATAATCATAAT 306

RESULT 100
AA264969
LOCUS       AA264969                324 bp    mRNA    linear   EST 19-APR-2001
DEFINITION  LD08623.5prime LD Drosophila melanogaster embryo BlueScript
            Drosophila melanogaster cDNA clone LD08623 5prime, mRNA sequence.
ACCESSION   AA264969
VERSION     AA264969.1  GI:1901031
KEYWORDS    EST.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 324)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Blast expect value = 1.4E-110 on X63731: D.melanogaster regulatory
            signal sequence (clone scs)
            Plate: 86  row: B  column: 11
            High quality sequence stop: 273.
FEATURES             Location/Qualifiers
     source          1. .324
                     /organism="Drosophila melanogaster"
                     /mol_type="mRNA"
                     /db_xref="BDGP_EST:BDcln007954"
                     /db_xref="taxon:7227"
                     /clone="LD08623"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="SOLR"
                     /clone_lib="LD Drosophila melanogaster embryo BlueScript"
                     /note="Organ: embryo; Vector: BlueScript SK; Site_1:
                     EcoRI; Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
                     Synthesis kit. Oligo dT-primed and directionally cloned at
                     EcoRI and XhoI in BlueScript SK(+/-)"
ORIGIN

Alignment Scores:
Pred. No.:                1.82e+03        Length:         324
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       1               Gaps:           0

US-09-858-332G-28 (1-6) x AA264969 (1-324)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        171 CACAACCATAATCACAAT 188

RESULT 101
AW664206
LOCUS       AW664206                324 bp    mRNA    linear   EST 06-APR-2000
DEFINITION  hi06c06.x1 NCI_CGAP_GU1 Homo sapiens cDNA clone IMAGE:2971498 3'
            similar to contains element LTR4 repetitive element ;, mRNA
            sequence.
ACCESSION   AW664206
VERSION     AW664206.1  GI:7456747
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 324)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life
            Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The
            I.M.A.G.E. Consortium DNA Sequencing by: Washington University
            Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            image.llnl.gov/image/html/iresources.shtml
            Seq primer: -40UP from Gibco
            High quality sequence stop: 323.
FEATURES             Location/Qualifiers
     source          1. .324
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2971498"
                     /tissue_type="2 pooled high-grade transitional cell
                     tumors"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_GU1"
                     /note="Organ: genitourinary tract; Vector: pCMV-SPORT6;
                     Site_1: SalI; Site_2: NotI; Cloned unidirectionally.
                     Primer: Oligo dT.  Library constructed by Life
                     Technologies."
ORIGIN
```

```
Alignment Scores:
Pred. No.:                1.82e+03        Length:         324
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       1               Gaps:           0

US-09-858-332G-28 (1-6) x AW664206 (1-324)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        200 CATAATCACAATCATAAT 217

RESULT 102
DT050599/c
LOCUS       DT050599                324 bp    mRNA    linear   EST 11-AUG-2005
DEFINITION  COT_EB_E08 Fiber and Ovule of Xu-142 Lambda Zap Express Library
            Gossypium hirsutum cDNA 3' similar to unknown, mRNA sequence.
ACCESSION   DT050599
VERSION     DT050599.1  GI:72276366
KEYWORDS    EST.
SOURCE      Gossypium hirsutum (upland cotton)
  ORGANISM  Gossypium hirsutum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE   1  (bases 1 to 324)
  AUTHORS   Gou,J.Y., Li,C.H., Lin,Z.P., Wang,L.J. and Chen,X.Y.
  TITLE     ESTs from developing cotton fiber and embryo (2005)
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Chen,X.Y.
            National laboratory of Plant Molecular Genetics
            Institute of Plant Physiology of Shanghai Institutes for Biological
            Sciences, The Chinese Acadamy of Sciences
            300 Fenglin Road, Shanghai, 200032, P.R.China
            Tel: 86-21-54924033
            Fax: 86-21-54924015
            Email: xychen@sibs.ac.cn
            Seq primer: T7 Primer.
FEATURES             Location/Qualifiers
     source          1. .324
                     /organism="Gossypium hirsutum"
                     /mol_type="mRNA"
                     /cultivar="Xu-142"
                     /db_xref="taxon:3635"
                     /dev_stage="-3 to 25 days post anthesis"
                     /clone_lib="Fiber and Ovule of Xu-142 Lambda Zap Express
                     Library"
                     /note="Organ: Fiber and Ovule"
ORIGIN

Alignment Scores:
Pred. No.:                1.82e+03        Length:         324
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       8               Gaps:           0

US-09-858-332G-28 (1-6) x DT050599 (1-324)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         55 CATAATCATAATCATAAT 38

RESULT 103
BM536214
LOCUS       BM536214                325 bp    mRNA    linear   EST 20-FEB-2002
DEFINITION  EST589236 tomato breaker fruit Lycopersicon esculentum cDNA clone
            cLEG71L12 5' end, mRNA sequence.
ACCESSION   BM536214
VERSION     BM536214.1  GI:18815451
KEYWORDS    EST.
SOURCE      Lycopersicon esculentum (Solanum lycopersicum)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 325)
  AUTHORS   Alcala,J., Vrebalov,J., White,R., Vision,T., Karamycheva,S.A.,
            Tsai,J., Bougri,O., Kirkness,E., Utterback,T., Van Aken,S.,
            Ronning,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and
            Giovannoni,J.
  TITLE     Generation of ESTs from tomato fruit tissue, breaker stage (2002)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html
            This clone is available through the Clemson University Genomics
            Institute
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .325
                     /organism="Lycopersicon esculentum"
                     /mol_type="mRNA"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cLEG71L12"
                     /tissue_type="Pericarp"
                     /dev_stage="breaker"
                     /lab_host="SOLR"
                     /clone_lib="tomato breaker fruit"
                     /note="Vector: pBluescriptSKmCUadapt; Site_1: EcoR1;
                     Site_2: XhoI; supplier: Boyce Thompson Institute;
                     sequencing: The Institute for Genomic Research.  Fruit
                     were harvested at the breaker stage (first sign of
                     lycopene accumulation on the blossom end of fruit).  Fruit
                     were cut in half and the seeds and locules were discarded
                     prior to freezing the pericarp."
ORIGIN

Alignment Scores:
Pred. No.:                1.83e+03        Length:         325
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       3               Gaps:           0

US-09-858-332G-28 (1-6) x BM536214 (1-325)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        211 CACAATCACAATCACAAT 228

RESULT 104
CF962054
LOCUS       CF962054                325 bp    mRNA    linear   EST 15-JAN-2004
DEFINITION  9000rsiced_12884.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF962054
VERSION     CF962054.1  GI:38477921
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 325)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
```

```
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 325
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .325
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="trefoil"
                     /clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA
                     library"
ORIGIN

Alignment Scores:
Pred. No.:                1.83e+03        Length:         325
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       7               Gaps:           0

US-09-858-332G-28 (1-6) x CF962054 (1-325)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        219 CACAACCACAACCACAAC 236

RESULT 105
BJ437953/c
LOCUS       BJ437953                327 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ437953 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv35k23 3', mRNA sequence.
ACCESSION   BJ437953
VERSION     BJ437953.1  GI:19412675
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 327)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .327
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv35k23"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:                1.84e+03        Length:         327
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       3               Gaps:           0

US-09-858-332G-28 (1-6) x BJ437953 (1-327)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        256 CACAACCACAACCACAAC 239

RESULT 106
AW775956
LOCUS       AW775956                328 bp    mRNA    linear   EST 07-SEP-2000
DEFINITION  EST335021 DSIL Medicago truncatula cDNA clone pDSIL-3J5, mRNA
            sequence.
ACCESSION   AW775956
VERSION     AW775956.1  GI:7765769
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 328)
  AUTHORS   Fedorova,M., Pierson,B.L., Samac,D.A., Vance,C.P., Gantt,G.S.,
            Peng,H., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B.,
            Hansen,T.S., Holt,I.E. and Fraser,C.M.
  TITLE     ESTs from leaves of Medicago truncatula after inoculation with
            Colletotrichum trifolii
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Deborah A. Samac
            Department of Plant Pathology
            University of Minnesota
            495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA
            Tel: 612 625 1243
            Fax: 651 649 5058
            Email: debbys@puccini.crl.umn.edu
            Minnesota sequence name:M259289e
            TIGR sequence name:MTFAK51TK
            More information is available at. .
            http://chrysie.tamu.edu/medicago
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES             Location/Qualifiers
     source          1. .328
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /cultivar="genotype A17"
                     /db_xref="taxon:3880"
                     /clone="pDSIL-3J5"
                     /tissue_type="leaves infected with Colletotrichum
                     trifolii"
                     /dev_stage="cotyledons and primary leaves harvested 5 and
                     8 days after inoculation with Colletotrichum trifolii"
                     /lab_host="E. coli strain XLOLR"
                     /clone_lib="DSIL"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; cDNA was prepared from polyA+ enriched RNA from
                     cotyledons and primary leaves harvested 5 and 8 days after
                     inoculation with Colletotrichum trifolii. The cDNA was
                     directionally ligated into the Uni-ZAP XR vector from
                     Stratagene and packaged using Gigapack III Gold packaging
                     extracts. Plasmids containing cDNA inserts were excised
                     from the recombinant lambda-ZAP phage using Ex-Assist
```

helper phage and propagated in XLOLR cells. Note: EST may be of fungal origin."
ORIGIN

Alignment Scores:

Pred. No.:	1.85e+03	Length:	328
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0

US-09-858-332G-28 (1-6) x AW775956 (1-328)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         79 CATAATCATAATCATAAT 96
```

RESULT 107
BE591060
LOCUS BE591060 328 bp mRNA linear EST 18-AUG-2000
DEFINITION WHE0861_F08_L15ZS Wheat 20-45 DAP spike cDNA library Triticum aestivum cDNA clone WHE0861_F08_L15, mRNA sequence.
ACCESSION BE591060
VERSION BE591060.1 GI:9846054
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 328)
AUTHORS Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L. and Tong,J.C.
TITLE The structure and function of the expressed portion of the wheat genomes - 20-45 DAP spike cDNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oandersn@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
Seq primer: Stratagene SK primer.
FEATURES Location/Qualifiers
source 1. .328
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE0861_F08_L15"
/tissue_type="Spike and seed"
/dev_stage="Adult plant"
/lab_host="E. coli SOLR"
/clone_lib="Wheat 20-45 DAP spike cDNA library"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site_1: EcoRI; Site_2: XhoI; Plants were grown in the greenhouse. Spikes at 20 DAP and seeds at 30 to 45 DAP were harvested, total RNA and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give pBluescript phagemids in the TJ Close lab (Choi, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."
ORIGIN

Alignment Scores:

Pred. No.:	1.85e+03	Length:	328
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-28 (1-6) x BE591060 (1-328)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        308 CACAACCACAACCACAAT 325
```

RESULT 108
BX002884
LOCUS BX002884 329 bp DNA linear GSS 02-APR-2004
DEFINITION Arabidopsis thaliana T-DNA flanking sequence GK-367G12-017046, genomic survey sequence.
ACCESSION BX002884
VERSION BX002884.1 GI:26187844
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Li,Y., Rosso,M.G., Strizhov,N., Viehoever,P. and Weisshaar,B.
TITLE GABI-Kat SimpleSearch: a flanking sequence tag (FST) database for the identification of T-DNA insertion mutants in Arabidopsis thaliana
JOURNAL Bioinformatics 19 (11), 1441-1442 (2003)
PUBMED 12874060
REFERENCE 2
AUTHORS Rosso,M.G., Li,Y., Strizhov,N., Reiss,B., Dekker,K. and Weisshaar,B.
TITLE An Arabidopsis thaliana T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics
JOURNAL Plant Mol. Biol. 53 (1-2), 247-259 (2003)
PUBMED 14756321
REFERENCE 3
AUTHORS Strizhov,N., Li,Y., Rosso,M.G., Viehoever,P., Dekker,K.A. and Weisshaar,B.
TITLE High-throughput generation of sequence indexes from T-DNA mutagenized Arabidopsis thaliana lines
JOURNAL BioTechniques 35 (6), 1164-1168 (2003)
PUBMED 14682050
REFERENCE 4 (bases 1 to 329)
AUTHORS Li,Y., Rosso,M.G., Strizhov,N. and Weisshaar,B.
TITLE Direct Submission
JOURNAL Submitted (31-MAR-2004) Weisshaar B., Max-Planck-Institut fuer Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
COMMENT This sequence has been recovered from the left border of the T-DNA. It indicates an insertion within the locus defined by BAC clone f17h15. Details on the protocols used for generation of the sequence are described in References 1-3. The sequences are generated at the MPI for Plant Breeding Research in the context of the GABI-Kat project. GABI-Kat is part of the German Plant Genomics program designated 'GABI'. Information on line availability can be found at: http://www.mpiz-koeln.mpg.de/GABI-Kat/.
FEATURES Location/Qualifiers
source 1. .329
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/clone="GK-367G12-017046"
/clone_lib="Arabidopsis thaliana T-DNA insertion lines"
/ecotype="Col-0"
/note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from vector pAC161 (GenBank accession number: AJ537514). The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion.

T-DNA derived sequences were removed."
ORIGIN

Alignment Scores:

Pred. No.:	1.85e+03	Length:	329
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	10	Gaps:	0

US-09-858-332G-28 (1-6) x BX002884 (1-329)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         50 CATAATCATAACCATAAC 67
```

RESULT 109
CV504055
LOCUS CV504055 330 bp mRNA linear EST 04-OCT-2004
DEFINITION 70570.1 Mixed Floral Solanum tuberosum cDNA clone 70570 5', mRNA sequence.
ACCESSION CV504055
VERSION CV504055.1 GI:53786412
KEYWORDS EST.
SOURCE Solanum tuberosum (potato)
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; lamiids; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 330)
AUTHORS Flinn,B., Rothwell,C., Sardana,R., Griffiths,R., Lague,M., De Koeyer,D., Audy,P., Goyer,C., Li,X-Q., Wang-Pruski,G. and Regan,S.
TITLE Generation of ESTs from a normalized developing floral bud library
JOURNAL Unpublished (2004)
COMMENT Contact: Barry Flinn
The Canadian Potato Genome Project - BioAtlantech
921 College Hill Rd, Fredericton, ON, E3B 6Z9, CANADA
Email: bflinn@bioatlantech.nb.ca
Seq primer: T3.
FEATURES Location/Qualifiers
source 1. .330
/organism="Solanum tuberosum"
/mol_type="mRNA"
/cultivar="Shepody"
/db_xref="taxon:4113"
/clone="70570"
/tissue_type="Floral buds"
/lab_host="XL10-Gold"
/clone_lib="Mixed Floral"
/note="Vector: pBluescript II SK(+) XR; Site_1: EcoRI; Site_2: XhoI; supplier: Developmental series. Plants from pathogen-free Solanum tuberosum var. Shepody, clone 1756, nuclear stock were grown in a screenhouse under natural conditions. Developing floral buds were harvested at various sizes up to opening and pooled. RNA was isolated and used for library construction. A normalized library was constructed following a modified protocol of Bonaldo et al. (1996. Genome Research 6: 791-806)."
ORIGIN

Alignment Scores:

Pred. No.:	1.86e+03	Length:	330
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	7	Gaps:	0

US-09-858-332G-28 (1-6) x CV504055 (1-330)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         66 CATAATCATAATCATAAT 83
```

RESULT 110
AV427149
LOCUS AV427149 332 bp mRNA linear EST 10-MAR-2004
DEFINITION AV427149 Lotus japonicus young plants (two-week old) Lotus corniculatus var. japonicus cDNA clone MWM077a09_r 5', mRNA sequence.
ACCESSION AV427149
VERSION AV427149.1 GI:7786803
KEYWORDS EST.
SOURCE Lotus corniculatus var. japonicus (Lotus japonicus)
ORGANISM Lotus corniculatus var. japonicus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae; Lotus.
REFERENCE 1 (bases 1 to 332)
AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
TITLE Generation of 7137 non-redundant expressed sequence tags from a legume, Lotus japonicus
JOURNAL DNA Res. 7 (2), 127-130 (2000)
PUBMED 10819328
COMMENT Contact: Yasukazu Nakamura
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: ynakamu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES Location/Qualifiers
source 1. .332
/organism="Lotus corniculatus var. japonicus"
/mol_type="mRNA"
/isolate="Miyakojima MG-20"
/db_xref="taxon:34305"
/clone="MWM077a09_r"
/tissue_type="whole plant"
/dev_stage="two-week old plant"
/clone_lib="Lotus japonicus young plants (two-week old)"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI; synonym: Lotus japonicus"
ORIGIN

Alignment Scores:

Pred. No.:	1.87e+03	Length:	332
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0

US-09-858-332G-28 (1-6) x AV427149 (1-332)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db          4 CATAATCACAATCATAAT 21
```

RESULT 111
DT051762/c
LOCUS DT051762 333 bp mRNA linear EST 11-AUG-2005
DEFINITION COT_EP_H05 Fiber and Ovule of Xu-142 Lambda Zap Express Library Gossypium hirsutum cDNA 3' similar to unknown, mRNA sequence.
ACCESSION DT051762
VERSION DT051762.1 GI:72278896
KEYWORDS EST.
SOURCE Gossypium hirsutum (upland cotton)
ORGANISM Gossypium hirsutum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE 1 (bases 1 to 333)
AUTHORS Gou,J.Y., Li,C.H., Lin,Z.P., Wang,L.J. and Chen,X.Y.
TITLE ESTs from developing cotton fiber and embryo (2005)
JOURNAL Unpublished (2005)

```
COMMENT     Contact: Chen,X.Y.
            National laboratory of Plant Molecular Genetics
            Institute of Plant Physiology of Shanghai Institutes for Biological
            Sciences, The Chinese Acadamy of Sciences
            300 Fenglin Road, Shanghai, 200032, P.R.China
            Tel: 86-21-54924033
            Fax: 86-21-54924015
            Email: xychen@sibs.ac.cn
            Seq primer: T7 Primer.
FEATURES             Location/Qualifiers
     source          1. .333
                     /organism="Gossypium hirsutum"
                     /mol_type="mRNA"
                     /cultivar="Xu-142"
                     /db_xref="taxon:3635"
                     /dev_stage="-3 to 25 days post anthesis"
                     /clone_lib="Fiber and Ovule of Xu-142 Lambda Zap Express
                     Library"
                     /note="Organ: Fiber and Ovule"
ORIGIN

Alignment Scores:
Pred. No.:              1.88e+03        Length:         333
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     8               Gaps:           0

US-09-858-332G-28 (1-6) x DT051762 (1-333)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         43 CATAATCATAATCATAAT 26

RESULT 112
BF640271
LOCUS       BF640271                334 bp    mRNA    linear   EST 19-DEC-2000
DEFINITION  NF027C05IN1F1037 Insect herbivory Medicago truncatula cDNA clone
            NF027C05IN 5', mRNA sequence.
ACCESSION   BF640271
VERSION     BF640271.1  GI:11904429
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 334)
  AUTHORS   Korth,K., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,
            Flores,H.R., Inman,J.T., Weller,J.W. and May,G.D.
  TITLE     Expressed Sequence Tags from the Samuel Roberts Noble Foundation
            Medicago truncatula insect herbivory library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Korth K
            Dept. of Plant Pathology
            University of Arkansas
            217 Plant Science Building, Fayetteville, AR 72701, USA
            Tel: 501 575 5191
            Fax: 501 575 7601
            Email: kkorth@comp.uark.edu
            Insert Length: 334   Std Error: 0.00
            Plate: 027  row: C  column: 05
            Seq primer: TCACACAGGAAACAGCTATGAC.
FEATURES             Location/Qualifiers
     source          1. .334
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /db_xref="taxon:3880"
                     /clone="NF027C05IN"
                     /tissue_type="local and systemic leaves"
                     /dev_stage="mature"
                     /clone_lib="Insect herbivory"
                     /note="Vector: Lambda Zap; Library was produced from fully
                     expanded M. truncatula leaves of plants fed upon by
                     Spodoptera exigua (beet armyworm) for 24 hours. Systemic
                     (undamaged leaves from injured plants) and wounded leaves
                     were harvested and pooled."
ORIGIN

Alignment Scores:
Pred. No.:              1.88e+03        Length:         334
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-858-332G-28 (1-6) x BF640271 (1-334)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        155 CATAATCACAATCACAAT 172

RESULT 113
BG139303
LOCUS       BG139303                334 bp    mRNA    linear   EST 31-JAN-2001
DEFINITION  EST479745 wild tomato pollen Lycopersicon pennellii cDNA clone
            cLPP12B20 5' sequence, mRNA sequence.
ACCESSION   BG139303
VERSION     BG139303.1  GI:12639491
KEYWORDS    EST.
SOURCE      Lycopersicon pennellii (Solanum pennellii)
  ORGANISM  Lycopersicon pennellii
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 334)
  AUTHORS   van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Utterback,T.,
            Hansen,C., Ronning,C. and Tanksley,S.
  TITLE     Generation of ESTs from wild tomato (L. pennellii) pollen
  JOURNAL   Unpublished (2001)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html.
FEATURES             Location/Qualifiers
     source          1. .334
                     /organism="Lycopersicon pennellii"
                     /mol_type="mRNA"
                     /cultivar="TA56"
                     /db_xref="taxon:28526"
                     /clone="cLPP12B20"
                     /tissue_type="pollen"
                     /dev_stage="pollen collected from open flowers"
                     /lab_host="SOLR"
                     /clone_lib="wild tomato pollen"
                     /note="Vector: pBluescript SK(-); Site_1: EcoR1; Site_2:
                     Xho1; Pollen was collected from open flowers from
                     L.pennellii TA56, and stored at -80 C until library
                     construction."
ORIGIN

Alignment Scores:
Pred. No.:              1.88e+03        Length:         334
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-858-332G-28 (1-6) x BG139303 (1-334)

Qy          1 HisAsnHisAsnHisAsn 6
```

```
              ||||||||||||||||||
Db        230 CACAACCACAACCACAAC 247

RESULT 114
BE192611
LOCUS       BE192611                334 bp    mRNA    linear   EST 23-JUN-2000
DEFINITION  EST331340 R3TA Mus musculus cDNA clone R3TAD62, mRNA sequence.
ACCESSION   BE192611
VERSION     BE192611.1  GI:8687985
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 334)
  AUTHORS   Earle-Hughes,J., Quackenbush,J., Adams,M.D., Fraser,C.M. and
            Venter,J.C.
  TITLE     Generation of ESTs from Murine adipose, undifferentiated 3T3 cell
            line
  JOURNAL   Unpublished (2000)
COMMENT     Other_ESTs: TC86679
            Contact: Julie Earle-Hughes
            TIGR
            The Institute For Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208.
FEATURES             Location/Qualifiers
     source          1. .334
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /cultivar="3T3"
                     /db_xref="taxon:10090"
                     /clone="R3TAD62"
                     /tissue_type="adipose"
                     /dev_stage="undifferentiated 3T3 cell line"
                     /clone_lib="R3TA"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; Murine adipose, undifferentiated 3T3 cell line"
ORIGIN

Alignment Scores:
Pred. No.:              1.88e+03       Length:         334
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     2              Gaps:           0

US-09-858-332G-28 (1-6) x BE192611 (1-334)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        227 CACAACCATAACCACAAC 244

RESULT 115
BQ470935
LOCUS       BQ470935                334 bp    mRNA    linear   EST 30-MAY-2002
DEFINITION  HX04H16r HX Hordeum vulgare subsp. vulgare cDNA clone HX04H16
            5-PRIME, mRNA sequence.
ACCESSION   BQ470935
VERSION     BQ470935.1  GI:21278717
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 334)
  AUTHORS   Zhang,H., Weschke,W., Michalek,W., Stein,N. and Graner,A.
  TITLE     EST sequencing and analysis in barley (2002)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Stein Nils
            Molecular Markers Group, Department Genbank
            Institute of Plant Genetics and Crop Plant Research (IPK)
            Correnssstr. 3, 06466, Gatersleben, Germany
            Tel: 039482-5522
            Fax: 039482-5595
            Email: stein@ipk-gatersleben.de
            Insert Length: 334   Std Error: 0.00
            Plate: 4  row: H  column: 16
            Seq primer: M13rev.
FEATURES             Location/Qualifiers
     source          1. .334
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Barke"
                     /sub_species="vulgare"
                     /db_xref="taxon:112509"
                     /clone="HX04H16"
                     /tissue_type="apex"
                     /dev_stage="apex (3-5 mm in size)"
                     /lab_host="XL10-Gold"
                     /clone_lib="HX"
                     /note="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of
                     cDNA); Site_2: XhoI (3'-end of cDNA); Due to a cloning
                     artefact caused by the kit, in most cases the EcoRI site
                     is NOT present, as well as the EcoRIadapter used for
                     cloning. To excise the insert, restriction sites upstream
                     EcoRI should be used (e.g. BamHI, SalI,PstI). NOTE: Also
                     due to the cloning system used Blue/white selection for
                     recombinats is not 100% reliable."
ORIGIN

Alignment Scores:
Pred. No.:              1.88e+03       Length:         334
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     5              Gaps:           0

US-09-858-332G-28 (1-6) x BQ470935 (1-334)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        248 CACAACCACAACCACAAC 265

RESULT 116
BI234561
LOCUS       BI234561                335 bp    mRNA    linear   EST 12-JUL-2001
DEFINITION  RE30278.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
            Drosophila melanogaster cDNA clone RE30278 5 similar to CG14732:
            FBan0014732 located on: 3R 87A6-87A6;: 04/12/2001, mRNA sequence.
ACCESSION   BI234561
VERSION     BI234561.1  GI:14703004
KEYWORDS    EST.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 335)
  AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B.,
            Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E.,
            George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G.,
            Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
            Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and
            Rubin,G.M.
  TITLE     BDGP/HHMI RE Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
```

```
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003693: arm:3R [7544898,7771380]
            estimated-cyto:86F7-87A8: 04/12/2001
            Plate: RE.302  row: G  column: 6
            High quality sequence stop: 334.
FEATURES              Location/Qualifiers
     source           1. .335
                      /organism="Drosophila melanogaster"
                      /mol_type="mRNA"
                      /db_xref="taxon:7227"
                      /clone="RE30278"
                      /sex="male and female"
                      /dev_stage="0-24 hours mixed stage embryonic"
                      /lab_host="DH5-alpha TonA"
                      /clone_lib="RE Drosophila melanogaster normalized Embryo
                      pFlc-1"
                      /note="Organ: embryo; Vector: pFlc1; Site_1: XhoI; Site_2:
                      BamHI; Library was kindly generated by Piero Carninci at
                      the RIKEN. The library  was normalized and excised using
                      Cre recombinase. Plasmid cDNA library."
ORIGIN

Alignment Scores:
Pred. No.:                1.89e+03        Length:         335
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       2               Gaps:           0

US-09-858-332G-28 (1-6) x BI234561 (1-335)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        209 CACAACCATAATCACAAT 226

RESULT 117
AW127263
LOCUS       AW127263                336 bp    mRNA    linear   EST 07-SEP-2000
DEFINITION  M110204 GVN Medicago truncatula cDNA clone N244, mRNA sequence.
ACCESSION   AW127263
VERSION     AW127263.2  GI:7771326
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 336)
  AUTHORS   Fedorova,M., Pierson,B.L., Samac,D.A., Gantt,J.S., Vance,C.P.,
            Gonzales,M.B. and Ellis,L.
  TITLE     ESTs from effective root nodules of Medicago truncatula
  JOURNAL   Unpublished (2000)
COMMENT     On Oct 25, 1999 this sequence version replaced gi:6115167.
            Contact: Carroll P. Vance
            Department of Agronomy and Plant Genetics
            University of Minnesota
            411 Borlaug Hall, 1991 Upper Buford Circle, St.Paul, MN 55108 USA
            Tel: 612 625 5715
            Fax: 651-649-5058
            Email: vance004@maroon.tc.umn.edu
            Date: 9/8/99; Updated to the Database of Expressed Sequence Tags
            (dbEST) on 04/27/00; More information is available at
            'http://chrysie.tamu.edu/medicago'.
            Seq primer: T3.
FEATURES              Location/Qualifiers
     source           1. .336
                      /organism="Medicago truncatula"
                      /mol_type="mRNA"
                      /cultivar="genotype A17"
                      /db_xref="taxon:3880"
                      /clone="N244"
                      /tissue_type="N2-fixing root nodules"
                      /dev_stage="effective root nodules harvested one month
                      post inoculation with Sinorhizobium meliloti"
                      /lab_host="E. coli strain XLOLR"
                      /clone_lib="GVN"
                      /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                      XhoI; cDNA was prepared from polyA+ enriched RNA from
                      effective root nodules harvested one month post
                      inoculation with Sinorhizobium meliloti. The cDNA was
                      directionally ligated into the Uni-ZAP XR vector from
                      Stratagene and packaged using Gigapack III Gold packaging
                      extracts. Plasmids containing cDNA inserts were excised
                      from the recombinant lambda-ZAP phage using Ex-Assist
                      helper phage and propagated in XLOLR cells."
ORIGIN

Alignment Scores:
Pred. No.:                1.9e+03         Length:         336
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       1               Gaps:           0

US-09-858-332G-28 (1-6) x AW127263 (1-336)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        204 CACAATCACAATCACAAC 221

RESULT 118
CB211509
LOCUS       CB211509                336 bp    mRNA    linear   EST 05-FEB-2003
DEFINITION  OML01789 Oryza minuta HybriZAP-2.1 XR library Oryza minuta cDNA 5',
            mRNA sequence.
ACCESSION   CB211509
VERSION     CB211509.1  GI:28257600
KEYWORDS    EST.
SOURCE      Oryza minuta
  ORGANISM  Oryza minuta
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 336)
  AUTHORS   Shin,J.S.
  TITLE     Oryza minuta HybriZAP-2.1 XR library
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Jeong Sheop Shin
            Plant Molecular Genetics
            Graduate School of Biotechnology, University of Korea
            136-701 Anam-dong 5/1 Seoul, Korea
            Tel: 00 82 2 3290 3430
            Fax: 00 82 2 927 9028
            Email: jsshin@kuccnx.korea.ac.kr.
FEATURES              Location/Qualifiers
     source           1. .336
                      /organism="Oryza minuta"
                      /mol_type="mRNA"
                      /db_xref="taxon:63629"
                      /dev_stage="4-weeks after germination"
                      /clone_lib="Oryza minuta HybriZAP-2.1 XR library"
                      /note="Organ: immature leaf"
ORIGIN

Alignment Scores:
Pred. No.:                1.9e+03         Length:         336
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0
```

```
US-09-858-332G-28 (1-6) x CB211509 (1-336)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        209 CACAACCACAACCACAAC 226

RESULT 119
C93386
LOCUS       C93386                   337 bp    mRNA    linear   EST 12-JUL-1999
DEFINITION  C93386 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
            discoideum cDNA clone SSM778, mRNA sequence.
ACCESSION   C93386
VERSION     C93386.1  GI:3075262
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 337)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     The Dictyostelium developmental cDNA project: generation and
            analysis of expressed sequence tags from the first-finger stage of
            development
  JOURNAL   DNA Res. 5 (6), 335-340 (1998)
   PUBMED   10048482
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp
            PROJECT = 'Dictyostelium discoideum cDNA project in Japan'
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .337
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SSM778"
                     /dev_stage="slug"
                     /clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
ORIGIN

Alignment Scores:
Pred. No.:              1.9e+03        Length:         337
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     5              Gaps:           0

US-09-858-332G-28 (1-6) x C93386 (1-337)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         38 CACAACCACAACCACAAC 55

RESULT 120
CB212482
LOCUS       CB212482                 337 bp    mRNA    linear   EST 05-FEB-2003
DEFINITION  OML02762 Oryza minuta HybriZAP-2.1 XR library Oryza minuta cDNA 5',
            mRNA sequence.
ACCESSION   CB212482
VERSION     CB212482.1  GI:28258573
KEYWORDS    EST.
SOURCE      Oryza minuta
  ORGANISM  Oryza minuta
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 337)
  AUTHORS   Shin,J.S.
  TITLE     Oryza minuta HybriZAP-2.1 XR library
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Jeong Sheop Shin
            Plant Molecular Genetics
            Graduate School of Biotechnology, University of Korea
            136-701 Anam-dong 5/1 Seoul, Korea
            Tel: 00 82 2 3290 3430
            Fax: 00 82 2 927 9028
            Email: jsshin@kuccnx.korea.ac.kr.
FEATURES             Location/Qualifiers
     source          1. .337
                     /organism="Oryza minuta"
                     /mol_type="mRNA"
                     /db_xref="taxon:63629"
                     /dev_stage="4-weeks after germination"
                     /clone_lib="Oryza minuta HybriZAP-2.1 XR library"
                     /note="Organ: immature leaf"
ORIGIN

Alignment Scores:
Pred. No.:              1.9e+03        Length:         337
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     6              Gaps:           0

US-09-858-332G-28 (1-6) x CB212482 (1-337)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        210 CACAACCACAACCACAAC 227

RESULT 121
CF369337/c
LOCUS       CF369337                 337 bp    mRNA    linear   EST 27-AUG-2003
DEFINITION  rg58e09.y1 Meloidogyne hapla female SMART pGEM Meloidogyne hapla
            cDNA 5', mRNA sequence.
ACCESSION   CF369337
VERSION     CF369337.1  GI:34316583
KEYWORDS    EST.
SOURCE      Meloidogyne hapla
  ORGANISM  Meloidogyne hapla
            Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
            Tylenchoidea; Meloidogynidae; Meloidogyninae; Meloidogyne.
REFERENCE   1  (bases 1 to 337)
  AUTHORS   McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J.,
            Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B.,
            Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C.,
            Tsagareishvili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C.,
            Underwood,K., Steptoe,M., Allen,M., Person,B., Swaller,T.,
            Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     The Washington Univ. Nematode EST Project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: McCarter JP
            The Washington Univ. Nematode EST Project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Cloned unidirectionally. Poly(A)+ RNA was concentrated and purified
            using Dynabeads (Dynal) and mRNA eluted for first strand synthesis.
            First strand cDNA was created using MMLV RT (Powerscript, Clontech)
            and primed with oligo(dT) with XhoI site and 5'SMART 'anchor' added
            using chimeric DNA-RNA oligo. 12 PCR cycles were done using first
            strand and primers specific to SMART oligo and 3' end. Double
            stranded cDNA was digested using XhoI/NotI, fractioned on
            Chroma-spin 400 columns (Clontech) and ligated to digested
            pGEM-11zf(+) plasmid. Chemically competent DH10B cells were used as
```

host cells. Library materials provided by Dr. David Bird of North Carolina State University. Library construction by Jeff Rousch. See www.nematode.net for additional project information.
Seq primer: Sp6.

FEATURES Location/Qualifiers
source 1. .337
/organism="Meloidogyne hapla"
/mol_type="mRNA"
/db_xref="taxon:6305"
/tissue_type="whole organism"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="Meloidogyne hapla female SMART pGEM"
/note="Vector: plasmid (ampicillin resistant); Site_1: XhoI; Site_2: NotI; Cloned unidirectionally. Poly(A)+ RNA was concentrated and purified using Dynabeads (Dynal) and mRNA eluted for first strand synthesis. First strand cDNA was created using MMLV RT (Powerscript, Clontech) and primed with oligo(dT) with XhoI site and 5'SMART 'anchor' added using chimeric DNA-RNA oligo. 12 PCR cycles were done using first strand and primers specific to SMART oligo and 3' end. Double stranded cDNA was digested using XhoI/NotI, fractioned on Chroma-spin 400 columns (Clontech) and ligated to digested pGEM-11zf(+) plasmid. Chemically competent DH10B cells were used as host cells. Library materials provided by Dr. David Bird of North Carolina State University. Library construction by Jeff Rousch. See www.nematode.net for additional project information."

ORIGIN

Alignment Scores:

Pred. No.:	1.9e+03	Length:	337
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-09-858-332G-28 (1-6) x CF369337 (1-337)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        204 CATAATCACAATCATAAT 187
```

RESULT 122
BF293983
LOCUS BF293983 338 bp mRNA linear EST 17-NOV-2000
DEFINITION WHE2162_D08_H16ZS Triticum turgidum L. var. durum (durum wheat) whole plant cDNA library Triticum turgidum cDNA clone WHE2162_D08_H16, mRNA sequence.
ACCESSION BF293983
VERSION BF293983.1 GI:11225047
KEYWORDS EST.
SOURCE Triticum turgidum
ORGANISM Triticum turgidum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 338)
AUTHORS Akhunov,E., Anderson,O.D., Chao,S., Chin,A., Choi,D.W., Close,T.J., Fenton,R.D., Han,P.S., Hsia,C.C., Kang,Y., Kianian,P., Lazo,G.R., Miller,R., Otto,C., Rausch,C.J., Seaton,C.L., Simons,K., Tong,J.C. and Zhang,D.
TITLE The structure and function of the expressed portion of the wheat genomes - Whole plant cDNA library from Triticum turgidum L. var. durum
JOURNAL Unpublished (2000)
COMMENT Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oandersn@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
Seq primer: Stratagene SK primer.

FEATURES Location/Qualifiers
source 1. .338
/organism="Triticum turgidum"
/mol_type="mRNA"
/cultivar="Langdon-16"
/db_xref="taxon:4571"
/clone="WHE2162_D08_H16"
/tissue_type="All tissues"
/dev_stage="Different growth stages"
/lab_host="E. coli SOLR"
/clone_lib="Triticum turgidum L. var. durum (durum wheat) whole plant cDNA library"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site_1: EcoRI; Site_2: XhoI; Plants were grown in a growth chamber at North Dakota State University (Kianian, Otto, Simons). Tissues collected from seven-day etiolated seedling leaf, stem, root and seed; leaf from plant at fourth leaf stage; spike from pre-anthesis through 20 days after anthesis; flag leaf; leaf and stem tissue from tillers. and root. Total RNA and poly(A) RNA were prepared from each tissue and then pooled, a cDNA library was made, and the cDNA clones were in vivo excised to give pBluescript phagemids in the TJ Close lab (Akhunov, Chin, Choi, Close, Fenton, Kianian, Otto, Simons, Zhang) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

ORIGIN

Alignment Scores:

Pred. No.:	1.91e+03	Length:	338
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-28 (1-6) x BF293983 (1-338)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        278 CACAACCACAACCACAAT 295
```

RESULT 123
BM536213
LOCUS BM536213 338 bp mRNA linear EST 20-FEB-2002
DEFINITION EST589235 tomato breaker fruit Lycopersicon esculentum cDNA clone cLEG71L10 5' end, mRNA sequence.
ACCESSION BM536213
VERSION BM536213.1 GI:18815450
KEYWORDS EST.
SOURCE Lycopersicon esculentum (Solanum lycopersicum)
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 338)
AUTHORS Alcala,J., Vrebalov,J., White,R., Vision,T., Karamycheva,S.A., Tsai,J., Bougri,O., Kirkness,E., Utterback,T., Van Aken,S., Ronning,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
TITLE Generation of ESTs from tomato fruit tissue, breaker stage (2002)
JOURNAL Unpublished (2002)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html

```
            This clone is available through the Clemson University Genomics
            Institute
            Seq primer: T3.
FEATURES              Location/Qualifiers
     source           1. .338
                      /organism="Lycopersicon esculentum"
                      /mol_type="mRNA"
                      /cultivar="TA496"
                      /db_xref="taxon:4081"
                      /clone="cLEG71L10"
                      /tissue_type="Pericarp"
                      /dev_stage="breaker"
                      /lab_host="SOLR"
                      /clone_lib="tomato breaker fruit"
                      /note="Vector: pBluescriptSKmCUadapt; Site_1: EcoR1;
                      Site_2: XhoI; supplier: Boyce Thompson Institute;
                      sequencing: The Institute for Genomic Research.  Fruit
                      were harvested at the breaker stage (first sign of
                      lycopene accumulation on the blossom end of fruit).  Fruit
                      were cut in half and the seeds and locules were discarded
                      prior to freezing the pericarp."
ORIGIN

Alignment Scores:
Pred. No.:                1.91e+03        Length:         338
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       3               Gaps:           0

US-09-858-332G-28 (1-6) x BM536213 (1-338)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        211 CACAATCACAATCACAAT 228

RESULT 124
BQ990783/c
LOCUS       BQ990783                338 bp    mRNA    linear   EST 21-AUG-2002
DEFINITION  QGF21B13.yg.ab1 QG_EFGHJ lettuce serriola Lactuca sativa cDNA clone
            QGF21B13, mRNA sequence.
ACCESSION   BQ990783
VERSION     BQ990783.1  GI:22410318
KEYWORDS    EST.
SOURCE      Lactuca sativa
  ORGANISM  Lactuca sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
            Cichorieae; Lactuca.
REFERENCE   1  (bases 1 to 338)
  AUTHORS   Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J.,
            Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y.,
            Lai,Z., Church,S., Jackson,L. and Bradford,K.
  TITLE     Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
            belongs to contig QG_CA_Contig3631, see http://cgpdb.ucdavis.edu/
            for details.
            Plate: QGF21  row: B  column: 13.
FEATURES              Location/Qualifiers
     source           1. .338
                      /organism="Lactuca sativa"
                      /mol_type="mRNA"
                      /cultivar="L.serriola"
                      /db_xref="taxon:4236"
                      /clone="QGF21B13"
                      /lab_host="E.coli"
                      /clone_lib="QG_EFGHJ lettuce serriola"
                      /note="Vector: pBRcDNASfiAB; The library was constructed
                      from 10 different sources of RNA from a single genotype.
                      Separate cDNAs were generated using primers that
                      incorporated unique 5' and 3' tags to distinguish each
                      source of RNA. cDNAs were then pooled, size-fractionated,
                      directionally cloned into a custom medium-copy vector and
                      transformations made with four size classes to minimize
                      size bias. Details of each source of RNA and library
                      construction can be obtained at http://cgpdb.ucdavis.edu/
                      TAG_TISSUE=flowers environmental stress
                      TAG_LIB=QG_EFGHJ lettuce serriola
                      TAG_SEQ=CGAATGCGGG"
ORIGIN

Alignment Scores:
Pred. No.:                1.91e+03        Length:         338
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       5               Gaps:           0

US-09-858-332G-28 (1-6) x BQ990783 (1-338)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        197 CATAATCATAATCATAAT 180

RESULT 125
CF369304/c
LOCUS       CF369304                338 bp    mRNA    linear   EST 27-AUG-2003
DEFINITION  rg58b07.y1 Meloidogyne hapla female SMART pGEM Meloidogyne hapla
            cDNA 5', mRNA sequence.
ACCESSION   CF369304
VERSION     CF369304.1  GI:34316550
KEYWORDS    EST.
SOURCE      Meloidogyne hapla
  ORGANISM  Meloidogyne hapla
            Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
            Tylenchoidea; Meloidogynidae; Meloidogyninae; Meloidogyne.
REFERENCE   1  (bases 1 to 338)
  AUTHORS   McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J.,
            Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B.,
            Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C.,
            Tsagareishvili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C.,
            Underwood,K., Steptoe,M., Allen,M., Person,B., Swaller,T.,
            Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     The Washington Univ. Nematode EST Project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: McCarter JP
            The Washington Univ. Nematode EST Project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Cloned unidirectionally. Poly(A)+ RNA was concentrated and purified
            using Dynabeads (Dynal) and mRNA eluted for first strand synthesis.
            First strand cDNA was created using MMLV RT (Powerscript, Clontech)
            and primed with oligo(dT) with XhoI site and 5'SMART 'anchor' added
            using chimeric DNA-RNA oligo. 12 PCR cycles were done using first
            strand and primers specific to SMART oligo and 3' end. Double
            stranded cDNA was digested using XhoI/NotI, fractioned on
            Chroma-spin 400 columns (Clontech) and ligated to digested
            pGEM-11zf(+) plasmid. Chemically competent DH10B cells were used as
            host cells. Library materials provided by Dr. David Bird of North
            Carolina State University. Library construction by Jeff Rousch. See
```

```
          www.nematode.net for additional project information.
          Seq primer: Sp6.
FEATURES             Location/Qualifiers
     source          1. .338
                     /organism="Meloidogyne hapla"
                     /mol_type="mRNA"
                     /db_xref="taxon:6305"
                     /tissue_type="whole organism"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="Meloidogyne hapla female SMART pGEM"
                     /note="Vector: plasmid (ampicillin resistant); Site_1:
                     XhoI; Site_2: NotI; Cloned unidirectionally. Poly(A)+ RNA
                     was concentrated and purified using Dynabeads (Dynal) and
                     mRNA eluted for first strand synthesis. First strand cDNA
                     was created using MMLV RT (Powerscript, Clontech) and
                     primed with oligo(dT) with XhoI site and 5'SMART 'anchor'
                     added using chimeric DNA-RNA oligo. 12 PCR cycles were
                     done using first strand and primers specific to SMART
                     oligo and 3' end. Double stranded cDNA was digested using
                     XhoI/NotI, fractioned on Chroma-spin 400 columns
                     (Clontech) and ligated to digested pGEM-11zf(+) plasmid.
                     Chemically competent DH10B cells were used as host cells.
                     Library materials provided by Dr. David Bird of North
                     Carolina State University. Library construction by Jeff
                     Rousch. See www.nematode.net for additional project
                     information."
ORIGIN

Alignment Scores:
Pred. No.:                1.91e+03        Length:         338
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x CF369304 (1-338)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        205 CATAATCACAATCATAAT 188

RESULT 126
CN926847
LOCUS       CN926847                338 bp    mRNA    linear   EST 07-JUN-2004
DEFINITION  000502AEPA002318HT (AEPA) Pinkie expanding leaf Malus x domestica
            cDNA clone AEPA002318, mRNA sequence.
ACCESSION   CN926847
VERSION     CN926847.1  GI:48399660
KEYWORDS    EST.
SOURCE      Malus x domestica
  ORGANISM  Malus x domestica
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Rosales; Rosaceae; Maloideae; Malus.
REFERENCE   1  (bases 1 to 338)
  AUTHORS   Beuning,L., Bowen,J., Crowhurst,R., Gleave,A., Janssen,B.,
            McArtney,S., Newcomb,R., Ross,G., Snowden,K., Walton,E. and Yauk,Y.
  TITLE     HortResearch Apple EST Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Gleave,A.
            Sequencing Facility
            The Horticulture and Food Research Institute of New Zealand Ltd
            120 Mt Albert Rd, Mt Albert, Auckland, New Zealand
            Tel: 00 64 09 815 4200
            Fax: 00 64 09 815 4201
            Email: est@hortresearch.co.nz.
FEATURES             Location/Qualifiers
     source          1. .338
                     /organism="Malus x domestica"
                     /mol_type="mRNA"
                     /db_xref="taxon:3750"
                     /clone="AEPA002318"
                     /tissue_type="Leaf"
                     /dev_stage="Expanding"
                     /clone_lib="(AEPA) Pinkie expanding leaf"
                     /note="Vector: pBK-CMV; Library sequenced by Genesis
                     Research & Development"
ORIGIN

Alignment Scores:
Pred. No.:                1.91e+03        Length:         338
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       7               Gaps:           0

US-09-858-332G-28 (1-6) x CN926847 (1-338)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        185 CACAACCACAACCACAAC 202

RESULT 127
CG898233/c
LOCUS       CG898233                338 bp    DNA     linear   GSS 08-DEC-2003
DEFINITION  pastbac066xb17.b1.ab1 Res147 1 Pasteuria penetrans genomic 5',
            genomic survey sequence.
ACCESSION   CG898233
VERSION     CG898233.1  GI:39553728
KEYWORDS    GSS.
SOURCE      Pasteuria penetrans
  ORGANISM  Pasteuria penetrans
            Bacteria; Firmicutes; Bacillales; Alicyclobacillaceae; Pasteuria.
REFERENCE   1  (bases 1 to 338)
  AUTHORS   Opperman,C.H., Davies,K.G., Sosinski,B.R., Waterman,J. and Burke,M.
  TITLE     Unpublished Data
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Opperman CH
            Center for the Biology of Nematode Parasitism and Nematode
            Interactions Unit
            North Carolina State University and Rothamsted Research, Ltd.
            Box 7253, NCSU, Raleigh, NC 27606, USA
            Email: warthog@unity.ncsu.edu
            Homology: e-val = 1e-10. Description: None found.  row: b  column:
            17
            Class: shotgun
            High quality sequence stop: 338.
FEATURES             Location/Qualifiers
     source          1. .338
                     /organism="Pasteuria penetrans"
                     /mol_type="genomic DNA"
                     /strain="Res147"
                     /db_xref="taxon:86005"
                     /clone_lib="Res147 1"
ORIGIN

Alignment Scores:
Pred. No.:                1.91e+03        Length:         338
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       10              Gaps:           0

US-09-858-332G-28 (1-6) x CG898233 (1-338)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        272 CACAACCACAACCACAAC 255

RESULT 128
BX551988/c
LOCUS       BX551988                340 bp    mRNA    linear   EST 10-OCT-2003
```

```
DEFINITION  BX551988 Glossina morsitans morsitans adult infected gut Glossina
            morsitans morsitans cDNA clone Tse120h10_q1c, mRNA sequence.
ACCESSION   BX551988
VERSION     BX551988.1  GI:33376301
KEYWORDS    EST.
SOURCE      Glossina morsitans morsitans
  ORGANISM  Glossina morsitans morsitans
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Hippoboscoidea; Glossinidae; Glossina.
REFERENCE   1  (bases 1 to 340)
  AUTHORS   Lehane,M.J., Aksoy,S., Gibson,W., Kerhornou,A., Berriman,M.,
            Hamilton,J., Soares,M.B., Bonaldo,M.F., Lehane,S. and Hall,N.
  TITLE     Adult midgut expressed sequence tags from the tsetse fly Glossina
            morsitans morsitans and expression analysis of putative immune
            response genes
  JOURNAL   Genome Biol. 4 (10), R63 (2003)
   PUBMED   14519198
COMMENT     Contact: Hall N
            Pathogen Sequencing Unit
            The Sanger Institute The Wellcome Trust Genome Campus
            Hinxton, Cambridge, CB10 1SA, UK
            Request for clones, please contact: Mike Lehane
            Prof. M.J.Lehane
            School of Biological Sciences,
            University of Wales,
            Bangor LL57 2UW
            All clones with suffix q1c are reverse primer reads starting at 5'
            end of the cDNA all p1c reads are from
            the 3' end.
FEATURES             Location/Qualifiers
     source          1. .340
                     /organism="Glossina morsitans morsitans"
                     /mol_type="mRNA"
                     /sub_species="morsitans"
                     /db_xref="taxon:37546"
                     /clone="Tse120h10_q1c"
                     /tissue_type="adult infected gut"
                     /clone_lib="Glossina morsitans morsitans adult infected
                     gut"
                     /note="country: Zimbabwe; EST from adult gut infected with
                     T.brucei"
ORIGIN

Alignment Scores:
Pred. No.:              1.92e+03        Length:         340
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     5               Gaps:           0

US-09-858-332G-28 (1-6) x BX551988 (1-340)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        274 CATAATCATAATCATAAT 257

RESULT 129
BP077382
LOCUS       BP077382                341 bp    mRNA    linear   EST 27-AUG-2004
DEFINITION  BP077382 Lotus japonicus roots Lotus corniculatus var. japonicus
            cDNA clone MR018g07_f 3', mRNA sequence.
ACCESSION   BP077382
VERSION     BP077382.1  GI:45634043
KEYWORDS    EST.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1  (bases 1 to 341)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Characteristics of the Lotus japonicus gene repertoire deduced from
            large-scale expressed sequence tag (EST) analysis
  JOURNAL   Plant Mol. Biol. 54 (3), 405-414 (2004)
   PUBMED   15284495
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .341
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="mRNA"
                     /isolate="Miyakojima MG-20"
                     /db_xref="taxon:34305"
                     /clone="MR018g07_f"
                     /tissue_type="roots"
                     /clone_lib="Lotus japonicus roots"
ORIGIN

Alignment Scores:
Pred. No.:              1.93e+03        Length:         341
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-858-332G-28 (1-6) x BP077382 (1-341)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         37 CATAATCACAATCATAAT 54

RESULT 130
BY114663
LOCUS       BY114663                341 bp    mRNA    linear   EST 08-DEC-2002
DEFINITION  BY114663 RIKEN full-length enriched, 18 days embryo whole body Mus
            musculus cDNA clone L430031D19 5', mRNA sequence.
ACCESSION   BY114663
VERSION     BY114663.1  GI:26225764
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 341)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
            Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
            Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
            Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
            Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
            Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
            Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
            Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
            Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
            Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
            Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
            Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
            Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
            Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
            Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
            Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
            Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
            Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
            Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
            Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
            Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
            Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
            Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
```

Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
PUBMED 12466851
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1. .341
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="L430031D19"
/tissue_type="whole body"
/dev_stage="18 days embryo"
/clone_lib="RIKEN full-length enriched, 18 days embryo whole body"
ORIGIN

Alignment Scores:

Pred. No.:	1.93e+03	Length:	341
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	5	Gaps:	0

US-09-858-332G-28 (1-6) x BY114663 (1-341)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        259 CACAACCATAACCACAAC 276
```

RESULT 131
BI271692
LOCUS BI271692 342 bp mRNA linear EST 18-JUL-2001
DEFINITION NF011B01FL1F1012 Developing flower Medicago truncatula cDNA clone NF011B01FL 5', mRNA sequence.
ACCESSION BI271692
VERSION BI271692.1 GI:14880395
KEYWORDS EST.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE 1 (bases 1 to 342)
AUTHORS Torres-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J., Flores,H.R., Inman,J.T., Weller,J.W. and May,G.D.
TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation Medicago truncatula flower library
JOURNAL Unpublished (2001)
COMMENT Contact: May GD
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 224 6650
Fax: 580 224 6692
Email: gdmay@noble.org
Insert Length: 342 Std Error: 0.00
Plate: 011 row: B column: 01
Seq primer: TCACACAGGAAACAGCTATGAC.
FEATURES Location/Qualifiers
source 1. .342
/organism="Medicago truncatula"
/mol_type="mRNA"
/db_xref="taxon:3880"
/clone="NF011B01FL"
/tissue_type="Developing flowers"
/dev_stage="Developmentally pooled. Contains a mixture of very young, developing, fully-opened flowers and flowers in early transition into pods."
/clone_lib="Developing flower"
/note="Vector: Lambda Zap; cDNA was prepared from polyA+ enriched, pooled samples of equivalent amounts of total RNA from very young, developing, fully-opened flowers and flowers transitioning into pods. The cDNA was directionally ligated into the Uni-Zap XR vector (Stratagene) and packaged using the Gigapack III Gold packaging extracts. Phagemids containing cDNA inserts were in vivo excised from the recombinant Uni-ZAP XR vector using ExAssist helper phage and the E. coli strain XL1-Blue MRF' (Stratagene). Excised plasmids were plated using SOLR cells."
ORIGIN

Alignment Scores:

Pred. No.:	1.93e+03	Length:	342
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-28 (1-6) x BI271692 (1-342)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        100 CATAACCACAACCACAAC 117
```

RESULT 132
BQ854048
LOCUS BQ854048 342 bp mRNA linear EST 14-AUG-2002
DEFINITION QGB22D09.yg.ab1 QG_ABCDI lettuce salinas Lactuca sativa cDNA clone QGB22D09, mRNA sequence.
ACCESSION BQ854048
VERSION BQ854048.1 GI:22239513
KEYWORDS EST.
SOURCE Lactuca sativa
ORGANISM Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;

```
            asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
            Cichorieae; Lactuca.
REFERENCE   1  (bases 1 to 342)
  AUTHORS   Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J.,
            Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y.,
            Lai,Z., Church,S., Jackson,L. and Bradford,K.
  TITLE     Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
            belongs to contig QG_CA_Contig3158, see http://cgpdb.ucdavis.edu/
            for details.
            Plate: QGB22  row: D  column: 09.
FEATURES             Location/Qualifiers
     source          1. .342
                     /organism="Lactuca sativa"
                     /mol_type="mRNA"
                     /cultivar="Salinas"
                     /db_xref="taxon:4236"
                     /clone="QGB22D09"
                     /lab_host="E.coli"
                     /clone_lib="QG_ABCDI lettuce salinas"
                     /note="Vector: pBRcDNASfiAB; The library was constructed
                     from 10 different sources of RNA from a single genotype.
                     Separate cDNAs were generated using primers that
                     incorporated unique 5' and 3' tags to distinguish each
                     source of RNA. cDNAs were then pooled, size-fractionated,
                     directionally cloned into a custom medium-copy vector and
                     transformations made with four size classes to minimize
                     size bias. Details of each source of RNA and library
                     construction can be obtained at http://cgpdb.ucdavis.edu/
                     TAG_TISSUE=roots
                     TAG_LIB=QG_ABCDI lettuce salinas
                     TAG_SEQ=GTTGCACGGG"
ORIGIN

Alignment Scores:
Pred. No.:              1.93e+03        Length:         342
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     5               Gaps:           0

US-09-858-332G-28 (1-6) x BQ854048 (1-342)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         58 CACAACCACAACCACAAC 75

RESULT 133
AL373429
LOCUS       AL373429                343 bp    mRNA    linear   EST 03-AUG-2000
DEFINITION  MtBA57G07F1 MtBA Medicago truncatula cDNA clone MtBA57G07 T3, mRNA
            sequence.
ACCESSION   AL373429
VERSION     AL373429.1  GI:9673182
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 343)
  AUTHORS   Journet,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jaillon,O.,
            Niebel,A., Carreau,V., Chatagnier,O., Kahn,D.,
            Gianinazzi-Pearson,V. and Gamas,P.
  TITLE     Medicago truncatula ESTs from nitrogen-starved roots
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de
            Biologie Moleculaire des Relations Plantes-Microorganismes,
            CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :
            Mt-est@toulouse.inra.fr Website :
            http://sequence.toulouse.inra.fr/Mtruncatula.html).
FEATURES             Location/Qualifiers
     source          1. .343
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /cultivar="Jemalong"
                     /db_xref="taxon:3880"
                     /clone="MtBA57G07"
                     /tissue_type="root tips"
                     /dev_stage="harvested after 3 days of N-starvation"
                     /clone_lib="MtBA"
                     /note="Vector: pBluescript pSK; Site_1: EcoRI; Site_2:
                     XhoI; Plants were grown in an aeroponic chamber for 14
                     days on nitrogen-rich medium followed by 3 days on N-free
                     medium. RNA was extracted from root tips (1-3 cm). cDNA
                     was prepared from polyA+ enriched RNA. The cDNA was
                     directionally ligated into Uni-zapXR vector from
                     Stratagene and packaged using Gigapack Gold packaging
                     extracts. Plasmids containing cDNA inserts were
                     mass-excised from phage stocks using ExAssit helper phage
                     and propagated in SOLR cells. Clone ordering and
                     sequencing was performed by the Centre National de
                     Sequencage (Genoscope, Evry, France)."
ORIGIN

Alignment Scores:
Pred. No.:              1.94e+03        Length:         343
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AL373429 (1-343)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        292 CACAACCATAACCACAAC 309

RESULT 134
AG915691
LOCUS       AG915691                343 bp    DNA     linear   GSS 01-FEB-2005
DEFINITION  Drosophila auraria DNA, clone: DAB1-013C13.R.fa, genomic survey
            sequence.
ACCESSION   AG915691
VERSION     AG915691.1  GI:58437047
KEYWORDS    GSS.
SOURCE      Drosophila auraria
  ORGANISM  Drosophila auraria
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1
  AUTHORS   Hattori,M., Toyoda,A., Murakami,K., Kuroki,Y., Fujiyama,A.,
            Toshio,T.K. and Sakaki,Y.
  TITLE     BAC end sequences of Library DAB1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 343)
  AUTHORS   Hattori,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JAN-2005) Masahira Hattori, The Institute of Physical
```

```
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the BAC library DAB1
            For BAC library availability, please contact Masa-Toshi Yamamoto
            (yamamoto@kit.jp).
            Submitted (30-11-2004) by  Masahira Hattori,
            RIKEN, Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp,Tel:81-45-503-9111,
            Fax:81-45-503-9170)
            This work was done in collaboration with Yamamoto, M-T. Drosophila
            Genetic Resource Center
            Saga Ippongi-cho, Ukyo-ku, Kyoto 616-8354, Japan
            Tel: 81-75-873-2660 FAX: 81-75-861-0881
            PRIMERS
                 Sequencing : R
            LIBRARY
                 Vector     : pKS150
                 R.Site 1   : SacI
                 R.Site 2   : SacI.
FEATURES             Location/Qualifiers
     source          1. .343
                     /organism="Drosophila auraria"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:47315"
                     /clone="DAB1-013C13.R.fa"
                     /clone_lib="DAB1 Drosophila BAC library"
ORIGIN

Alignment Scores:
Pred. No.:                1.94e+03        Length:          343
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       10              Gaps:            0

US-09-858-332G-28 (1-6) x AG915691 (1-343)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         31 CACAATCACAATCACAAC 48

RESULT 135
BY104996
LOCUS       BY104996                 344 bp    mRNA    linear   EST 07-DEC-2002
DEFINITION  BY104996 RIKEN full-length enriched, 13 days embryo whole body Mus
            musculus cDNA clone K630158M18 5', mRNA sequence.
ACCESSION   BY104996
VERSION     BY104996.1  GI:26215613
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 344)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
            Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
            Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
            Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
            Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
            Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
            Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
            Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
            Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
            Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
            Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
            Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
            Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
            Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
            Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
            Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
            Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
            Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
            Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
            Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
            Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
            Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
            Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
            Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
            Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
            Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
   PUBMED   12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
            Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
            Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,
            Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,
            Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and
            Hayashizaki,Y. Direct Submission
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
             cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .344
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="K630158M18"
                     /tissue_type="whole body"
                     /dev_stage="13 days embryo"
                     /clone_lib="RIKEN full-length enriched, 13 days embryo
                     whole body"
ORIGIN

Alignment Scores:
Pred. No.:                1.94e+03        Length:          344
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       5               Gaps:            0

US-09-858-332G-28 (1-6) x BY104996 (1-344)

Qy          1 HisAsnHisAsnHisAsn 6
```

```
                  ||||||||||||||||||
Db        275 CACAACCATAACCACAAC 292

RESULT 136
BQ849201
LOCUS       BQ849201                345 bp    mRNA    linear   EST 14-AUG-2002
DEFINITION  QGA9D24.yg.ab1 QG_ABCDI lettuce salinas Lactuca sativa cDNA clone
            QGA9D24, mRNA sequence.
ACCESSION   BQ849201
VERSION     BQ849201.1  GI:22234670
KEYWORDS    EST.
SOURCE      Lactuca sativa
  ORGANISM  Lactuca sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
            Cichorieae; Lactuca.
REFERENCE   1  (bases 1 to 345)
  AUTHORS   Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J.,
            Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y.,
            Lai,Z., Church,S., Jackson,L. and Bradford,K.
  TITLE     Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
            belongs to contig QG_CA_Contig3158, see http://cgpdb.ucdavis.edu/
            for details.
            Plate: QGA9  row: D  column: 24.
FEATURES             Location/Qualifiers
     source          1. .345
                     /organism="Lactuca sativa"
                     /mol_type="mRNA"
                     /cultivar="Salinas"
                     /db_xref="taxon:4236"
                     /clone="QGA9D24"
                     /lab_host="E.coli"
                     /clone_lib="QG_ABCDI lettuce salinas"
                     /note="Vector: pBRcDNASfiAB; The library was constructed
                     from 10 different sources of RNA from a single genotype.
                     Separate cDNAs were generated using primers that
                     incorporated unique 5' and 3' tags to distinguish each
                     source of RNA. cDNAs were then pooled, size-fractionated,
                     directionally cloned into a custom medium-copy vector and
                     transformations made with four size classes to minimize
                     size bias. Details of each source of RNA and library
                     construction can be obtained at http://cgpdb.ucdavis.edu/
                     TAG_TISSUE=roots
                     TAG_LIB=QG_ABCDI lettuce salinas
                     TAG_SEQ=GTTGCACGGG"
ORIGIN

Alignment Scores:
Pred. No.:              1.95e+03        Length:         345
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     5               Gaps:           0

US-09-858-332G-28 (1-6) x BQ849201 (1-345)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         58 CACAACCACAACCACAAC 75

RESULT 137
AA186278
LOCUS       AA186278                346 bp    mRNA    linear   EST 10-JAN-1997
DEFINITION  T3492 MVAT4 bloodstream form of serodeme WRATat1.1 Trypanosoma
            brucei rhodesiense cDNA 5', mRNA sequence.
ACCESSION   AA186278
VERSION     AA186278.1  GI:1772721
KEYWORDS    EST.
SOURCE      Trypanosoma brucei rhodesiense
  ORGANISM  Trypanosoma brucei rhodesiense
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 346)
  AUTHORS   Djikeng,A., Donelson,J.E. and Majiwa,P.A.O.
  TITLE     Generation of expressed sequence tags as physical landmarks in the
            genome of Trypanosoma brucei
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Majiwa PAO
            Molecular Biology Unit
            International Livestock Research Institute
            P.O. Box 30709, Nairobi, Kenya
            Tel: 254-2 630743
            Fax: 254-2 631499
            Email: p.majiwa@cgnet.com
            Seq primer: T3 primer.
FEATURES             Location/Qualifiers
     source          1. .346
                     /organism="Trypanosoma brucei rhodesiense"
                     /mol_type="mRNA"
                     /sub_species="rhodesiense"
                     /db_xref="taxon:31286"
                     /clone_lib="MVAT4 bloodstream form of serodeme WRATat1.1"
                     /note="Vector: Lambda ZAP II (Stratagene); Site_1: EcorI;
                     Site_2: XhoI; The mRNA was purified from a cloned
                     population of bloodstream trypanosomes reexpressing the
                     MVAT4 metacyclic variant surface glycoprotein (VSG). A
                     unidirectional oligo dT-primed EcoRI/XhoI cDNA library was
                     constructed in lambda ZAP II (Stratagene)."
ORIGIN

Alignment Scores:
Pred. No.:              1.96e+03        Length:         346
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AA186278 (1-346)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        310 CACAACCACAACCACAAC 327

RESULT 138
BP072036
LOCUS       BP072036                346 bp    mRNA    linear   EST 26-AUG-2004
DEFINITION  BP072036 Lotus corniculatus var. japonicus nodulating root Lotus
            corniculatus var. japonicus cDNA clone GNf063c04 3', mRNA sequence.
ACCESSION   BP072036
VERSION     BP072036.1  GI:45628748
KEYWORDS    EST.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1  (bases 1 to 346)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Characteristics of the Lotus japonicus gene repertoire deduced from
            large-scale expressed sequence tag (EST) analysis
  JOURNAL   Plant Mol. Biol. 54 (3), 405-414 (2004)
   PUBMED   15284495
```

```
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .346
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="mRNA"
                     /isolate="Gifu B-129"
                     /db_xref="taxon:34305"
                     /clone="GNf063c04"
                     /tissue_type="nodulating root"
                     /clone_lib="Lotus corniculatus var. japonicus nodulating
                     root"
                     /note="root of hypernodulating mutant har1-3, rich in
                     nodule primordia"
ORIGIN

Alignment Scores:
Pred. No.:              1.96e+03        Length:         346
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-858-332G-28 (1-6) x BP072036 (1-346)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         12 CATAATCACAATCATAAT 29

RESULT 139
CO267388
LOCUS       CO267388                346 bp    mRNA    linear   EST 23-JUN-2004
DEFINITION  EK150015.5prime Exelixis FlyTag CK01 pCDNA-SK+ Drosophila
            melanogaster cDNA clone EK150015 5, mRNA sequence.
ACCESSION   CO267388
VERSION     CO267388.1  GI:49152973
KEYWORDS    EST.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 346)
  AUTHORS   Kopczynski,C., Platt,D., Campbell,J., Muzong,C., Laufer,A.,
            Peterson,E. and Swimmer,C.
  TITLE     Exelixis FlyTag EST Project CK01 Library
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Based upon one or more reads of this clone where vector sequence
            was present at both ends, this clone has been determined to contain
            contain a cDNA insert on the order of 600-1000 bases.
            Plate: EK.1500  row: B  column: 3
            High quality sequence stop: 265.
FEATURES             Location/Qualifiers
     source          1. .346
                     /organism="Drosophila melanogaster"
                     /mol_type="mRNA"
                     /db_xref="taxon:7227"
                     /clone="EK150015"
                     /clone_lib="Exelixis FlyTag CK01 pCDNA-SK+"
                     /note="Organ: mixed stage embryos, imaginal disks, and
                     adult heads; Vector: pCDNA-SK+; Site_1: NotI; Site_2:
                     XhoI; Random primed, normalized library from mixed stage
                     embryos, imaginal disks, and adult heads."
ORIGIN

Alignment Scores:
Pred. No.:              1.96e+03        Length:         346
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     7               Gaps:           0

US-09-858-332G-28 (1-6) x CO267388 (1-346)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         24 CACAACCATAACCACAAC 41

RESULT 140
AA098068
LOCUS       AA098068                347 bp    mRNA    linear   EST 25-OCT-1996
DEFINITION  mk18e01.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
            IMAGE:493272 5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0 KD
            PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION   AA098068
VERSION     AA098068.1  GI:1643622
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 347)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:296720
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 326.
FEATURES             Location/Qualifiers
     source          1. .347
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:493272"
                     /dev_stage="19.5 dpc total fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /clone_lib="Soares mouse p3NMF19.5"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. RNA was kindly provided by
                     Dr. Minoru Ko (Wayne State University)."
ORIGIN

Alignment Scores:
```

```
Pred. No.:              1.96e+03        Length:         347
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AA098068 (1-347)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        318 CACAACCATAACCACAAC 335

RESULT 141
CF196076
LOCUS       CF196076                348 bp    mRNA    linear   EST 31-JUL-2003
DEFINITION  Azbi2.3_E06 AZ-BI Oryza sativa (japonica cultivar-group) cDNA clone
            AZ-BI2.3E06 5', mRNA sequence.
ACCESSION   CF196076
VERSION     CF196076.1  GI:33380733
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 348)
  AUTHORS   Ventelon-Debout,M., Thi-Thu Huang,N., Wissocq,A., Berger,C.,
            Michele,L., Piegu,B., Cooke,R., Ghesquiere,A., Delseny,M. and
            Brugidou,C.
  TITLE     Metabolism and Photosynthesis gene expression is affected in
            response to rice yellow mottle virus infection in Oryza sativa
            indica and japonica culltivars
  JOURNAL   Mol. Genet. Genomics (2003) In press
COMMENT     Contact: Brugidou C
            IRD
            BP 64501, 34394 Montpellier cedex 5 France
            Tel: 33 4 67 41 67 39
            Fax: 33 4 67 41 61 81
            PCR PRimers
            FORWARD: 5'AATTAACCCTCACTAAAGGG3'
            BACKWARD: 5'CGGGATATCACTCAGCATAATG3'
            Plate: 2.3  row: E  column: 6.
FEATURES             Location/Qualifiers
     source          1. .348
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Azucena"
                     /db_xref="taxon:39947"
                     /clone="AZ-BI2.3E06"
                     /tissue_type="leaves"
                     /dev_stage="3 weeks-old plants"
                     /lab_host="SOLR strain (Stratagene)"
                     /clone_lib="AZ-BI"
                     /note="Vector: pBluescript SK (+/-); Site_1: Eco RI;
                     Site_2: Xho I; Sample name: AZ-BI; Plant growth place:
                     greenhouse; Soil conditions: standard; Sowing date:
                     28/02/2000; Harvesting date: 18/03/2000; Stress date:
                     15/03/2000; Cultivar: azucena; Description: Az-wounded
                     leaves cDNA library; experimental condition: wound;"
ORIGIN

Alignment Scores:
Pred. No.:              1.97e+03        Length:         348
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x CF196076 (1-348)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        193 CACAACCACAACCACAAC 210

RESULT 142
BY776640
LOCUS       BY776640                349 bp    mRNA    linear   EST 23-MAR-2004
DEFINITION  BY776640 RIKEN full-length enriched, 17.5 days embryo whole body
            Mus musculus cDNA clone L930107E08 5', mRNA sequence.
ACCESSION   BY776640
VERSION     BY776640.1  GI:39703278
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 349)
  AUTHORS   Carninci,P., Waki,K., Shiraki,T., Konno,H., Shibata,K., Itoh,M.,
            Aizawa,K., Arakawa,T., Ishii,Y., Sasaki,D., Bono,H., Kondo,S.,
            Sugahara,Y., Saito,R., Osato,N., Fukuda,S., Sato,K., Watahiki,A.,
            Hirozane-Kishikawa,T., Nakamura,M., Shibata,Y., Yasunishi,A.,
            Kikuchi,N., Yoshiki,A., Kusakabe,M., Gustincich,S., Beisel,K.,
            Pavan,W., Aidinis,V., Nakagawara,A., Held,W.A., Iwata,H., Kono,T.,
            Nakauchi,H., Lyons,P., Wells,C., Hume,D.A., Fagiolini,M.,
            Hensch,T.K., Brinkmeier,M., Camper,S., Hirota,J., Mombaerts,P.,
            Muramatsu,M., Okazaki,Y., Kawai,J. and Hayashizaki,Y.
  TITLE     Targeting a complex transcriptome: the construction of the mouse
            full-length cDNA encyclopedia
  JOURNAL   Genome Res. 13 (6B), 1273-1289 (2003)
  PUBMED    12819125
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
             Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
             Please visit our web site (http://genome.gsc.riken.jp/) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .349
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="L930107E08"
                     /tissue_type="whole body"
                     /dev_stage="17.5 days embryo"
                     /clone_lib="RIKEN full-length enriched, 17.5 days embryo
                     whole body"
ORIGIN

Alignment Scores:
Pred. No.:              1.97e+03        Length:         349
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     5               Gaps:           0

US-09-858-332G-28 (1-6) x BY776640 (1-349)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        301 CACAACCATAACCACAAC 318

RESULT 143
```

```
CV252723/c
LOCUS       CV252723                349 bp    mRNA    linear   EST 22-SEP-2004
DEFINITION  PX0015.B21_M04 PT-X-FL-A-1 Populus trichocarpa cDNA clone
            PX0015_M04 3', mRNA sequence.
ACCESSION   CV252723
VERSION     CV252723.1  GI:52505698
KEYWORDS    EST.
SOURCE      Populus trichocarpa (Populus balsamifera subsp. trichocarpa)
  ORGANISM  Populus trichocarpa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Malpighiales; Salicaceae; Saliceae; Populus.
REFERENCE   1  (bases 1 to 349)
  AUTHORS   Ralph,S., Cooper,D., Kolosova,N., Oddy,C., Butterfield,Y.,
            Kirkpatrick,R., Liu,J., Palmquist,D., Stott,J., Barber,S., Yang,G.,
            Babakaiff,R., Brown-John,M., Chand,S., Featherstone,R., Masson,A.,
            Mayo,M., Moran,J., Olson,T., Wong,D., Ritland,C.E., Siddiqui,A.,
            Holt,R., Jones,S., Marra,M., Ellis,B.E., Douglas,C., Ritland,K. and
            Bohlmann,J.
  TITLE     The poplar transcriptome: Analysis of expressed sequence tags from
            multiple cDNA libraries
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Joerg Bohlmann
            Genome BC forest genomics program
            University of British Columbia
            Michael Smith Laboratories, 6174 University Boulevard, Rm. 237,
            Vancouver, British Columbia, Canada, V6T 1Z3
            Tel: 1-604-822-0282
            Fax: 1-604-822-2114
            Email: bohlmann@msl.ubc.ca
            Plate: PX0015  row: M  column: 04
            High quality sequence stop: 349
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .349
                     /organism="Populus trichocarpa"
                     /mol_type="mRNA"
                     /cultivar="383-2499 (Nisqually-1)"
                     /db_xref="taxon:3694"
                     /clone="PX0015_M04"
                     /sex="Female"
                     /lab_host="E. coli DH10B T1 phage resistant cells"
                     /clone_lib="PT-X-FL-A-1"
                     /note="Vector: pBluescript II SK (+) XR; Site_1: SstI (5'
                     end of cDNA); Site_2: XhoI (3' end of cDNA); Outer xylem
                     from 8 year old trees harvested within the Boise Cascade
                     region of Washington state on May 15th, 2001. cDNA was
                     prepared from 20 micrograms of mRNA according to the
                     full-length cDNA library construction method described by
                     Carninci P. et al. (2000), Genome Research
                     10(10):1617-1630 and directionally ligated into the
                     pBluescript II SK (+) XR vector digested with SstI (5'
                     end) and XhoI (3'). Plasmid DNA was then transformed by
                     electroporation into DH10B cells (Invitrogen) for
                     propagation."
ORIGIN

Alignment Scores:
Pred. No.:                1.97e+03        Length:         349
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       7               Gaps:           0

US-09-858-332G-28 (1-6) x CV252723 (1-349)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        304 CACAACCACAACCACAAC 287

RESULT 144
CV546329
LOCUS       CV546329                352 bp    mRNA    linear   EST 22-OCT-2004
DEFINITION  SS2467.txt NaCl treated Chinese cabbage seedling library Brassica
            rapa subsp. pekinensis cDNA clone SS2467.txt 5', mRNA sequence.
ACCESSION   CV546329
VERSION     CV546329.1  GI:54420861
KEYWORDS    EST.
SOURCE      Brassica rapa subsp. pekinensis
  ORGANISM  Brassica rapa subsp. pekinensis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 352)
  AUTHORS   Kang,J.S., Ryu,S.H., Yang,K.A. and Lim,C.O.
  TITLE     Expressed Sequence Tags of NaCl treated Chinese cabbage seedlings
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Lim, C.O.
            Plant Molecular Biology & Biotechnology Research Centre
            Gyeongsang National University
            #900 Gazwa-dong, Jinju 660-701, Korea
            Tel: 82 55 751 6255
            Fax: 82 55 759 9363
            Email: colim@nongae.gsnu.ac.kr
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .352
                     /organism="Brassica rapa subsp. pekinensis"
                     /mol_type="mRNA"
                     /cultivar="Jangwon"
                     /sub_species="pekinensis"
                     /db_xref="taxon:51351"
                     /clone="SS2467.txt"
                     /dev_stage="seedling"
                     /lab_host="XL-1 Blue"
                     /clone_lib="NaCl treated Chinese cabbage seedling library"
                     /note="Vector: pSPOT1; Site_1: Sal I; Site_2: Not I"
ORIGIN

Alignment Scores:
Pred. No.:                1.99e+03        Length:         352
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       7               Gaps:           0

US-09-858-332G-28 (1-6) x CV546329 (1-352)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        172 CACAACCACAACCACAAC 189

RESULT 145
AW127340
LOCUS       AW127340                354 bp    mRNA    linear   EST 07-SEP-2000
DEFINITION  M110512 DSIL Medicago truncatula cDNA clone IL32, mRNA sequence.
ACCESSION   AW127340
VERSION     AW127340.2  GI:7772056
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 354)
  AUTHORS   Fedorova,M., Pierson,B.L., Samac,D.A., Gantt,J.S., Vance,C.P.,
            Gonzales,M.B. and Ellis,L.
  TITLE     ESTs from Medicago truncatula leaves infected with Colletotrichum
            trifolii
  JOURNAL   Unpublished (2000)
COMMENT     On Oct 25, 1999 this sequence version replaced gi:6115244.
            Contact: Carroll P. Vance
            Department of Agronomy and Plant Genetics
```

```
            University of Minnesota
            411 Borlaug Hall, 1991 Upper Buford Circle, St.Paul, MN 55108 USA
            Tel: 612 625 5715
            Fax: 651-649-5058
            Email: vance004@maroon.tc.umn.edu
            Date: 9/15/99; Updated to the Database of Expressed Sequence Tags
            (dbEST) on 04/27/00; More information is available at
            'http://chrysie.tamu.edu/medicago'.
            Seq primer: T3.
FEATURES              Location/Qualifiers
     source           1. .354
                      /organism="Medicago truncatula"
                      /mol_type="mRNA"
                      /cultivar="genotype A17"
                      /db_xref="taxon:3880"
                      /clone="IL32"
                      /tissue_type="leaves infected with Colletotrichum
                      trifolii"
                      /dev_stage="cotyledons and primary leaves harvested 5 and
                      8 days after inoculation with Colletotrichum trifolii"
                      /lab_host="E. coli strain XLOLR"
                      /clone_lib="DSIL"
                      /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                      XhoI; cDNA was prepared from polyA+ enriched RNA from
                      cotyledons and primary leaves harvested 5 and 8 days after
                      inoculation with Colletotrichum trifolii. The cDNA was
                      directionally ligated into the Uni-ZAP XR vector from
                      Stratagene and packaged using Gigapack III Gold packaging
                      extracts. Plasmids containing cDNA inserts were excised
                      from the recombinant lambda-ZAP phage using Ex-Assist
                      helper phage and propagated in XLOLR cells. Note: EST may
                      be of fungal origin."
ORIGIN

Alignment Scores:
Pred. No.:                2e+03          Length:         354
Score:                    42.00          Matches:        6
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       1              Gaps:           0

US-09-858-332G-28 (1-6) x AW127340 (1-354)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        177 CACAACCATAACCATAAC 194

RESULT 146
BE604986
LOCUS       BE604986                354 bp    mRNA    linear   EST 21-AUG-2000
DEFINITION  WHE1713-1716_G09_G09ZS Wheat heat stressed spike cDNA library
            Triticum aestivum cDNA clone WHE1713-1716_G09_G09, mRNA sequence.
ACCESSION   BE604986
VERSION     BE604986.1  GI:9862256
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 354)
  AUTHORS   Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D.,
            Han,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Malatrasi,M., Miller,R.,
            Nguyen,H.T., Rausch,C.J., Seaton,C.L., Tong,J.C. and Zhang,D.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Heat stressed spike cDNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene SK primer.
FEATURES              Location/Qualifiers
     source           1. .354
                      /organism="Triticum aestivum"
                      /mol_type="mRNA"
                      /cultivar="Chinese Spring"
                      /db_xref="taxon:4565"
                      /clone="WHE1713-1716_G09_G09"
                      /tissue_type="Whole spike"
                      /dev_stage="Spikes at 5, 10, 15 and 20 days after
                      anthesis"
                      /lab_host="E. coli SOLR"
                      /clone_lib="Wheat heat stressed spike cDNA library"
                      /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                      Site_1: EcoRI; Site_2: XhoI; Spikes at 5, 10, 15 and 20
                      days after anthesis were heat stressed under  two
                      conditions at Texas Tech University (D. Zhang in HT Nguyrn
                      lab):  (1) at 38 C for 4 hours and (2) 5 days of cyclic
                      treatment of 38 C  for 4 hours.  Total RNA and poly(A) RNA
                      were prepared, a cDNA library  was made, and the cDNA
                      clones were in vivo excised to give pBluescript  phagemids
                      in the TJ Close lab (Choi, Close, Fenton, Malatrasi) at
                      the  University of California, Riverside.  Plasmid DNA
                      preparations and  DNA sequencing were performed in the OD
                      Anderson lab (all other authors)."
ORIGIN

Alignment Scores:
Pred. No.:                2e+03          Length:         354
Score:                    42.00          Matches:        6
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       2              Gaps:           0

US-09-858-332G-28 (1-6) x BE604986 (1-354)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         19 CACAACCACAACCACAAT 36

RESULT 147
BE240393/c
LOCUS       BE240393                355 bp    mRNA    linear   EST 12-JUL-2000
DEFINITION  EST404442 MHRP- Medicago truncatula cDNA clone pMHRP-45A11, mRNA
            sequence.
ACCESSION   BE240393
VERSION     BE240393.1  GI:9056577
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 355)
  AUTHORS   Harrison,M.J., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S.,
            Holt,I.E., Cho,J. and Fraser,C.M.
  TITLE     ESTs from phosphate-starved roots of Medicago truncatula
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Maria J. Harrison
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73401, USA
            Tel: 580-223-5810
            Fax: 580-221-7380
            Email: mjharrison@noble.org
            The Samuel Noble Roberts Foundation: N266260e
            TIGR sequence name:MTHAQ06TK
```

```
            More information is available at. .
            http://chrysie.tamu.edu/medicago
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES              Location/Qualifiers
     source           1. .355
                      /organism="Medicago truncatula"
                      /mol_type="mRNA"
                      /cultivar="A17"
                      /db_xref="taxon:3880"
                      /clone="pMHRP-45A11"
                      /tissue_type="roots"
                      /dev_stage="phosphate-starved"
                      /lab_host="XLOLR"
                      /clone_lib="MHRP-"
                      /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                      XhoI; At the trifoliate stage, M. truncatula plants were
                      transplanted to phosphate-free sand and grown for a
                      further 30 days. During  this period, they were fertilized
                      twice weekly with 1/2 Hoaglands  solutions containing 20uM
                      potassium phosphate. cDNA was prepared from polyA+
                      enriched RNA. The cDNA was directionally ligated into the
                      Unizap XR vector from Stratagene and packaged using
                      Gigapack III Gold packaging extracts. Plasmids containing
                      cDNA inserts  were excised from the recombinant lambda-Zap
                      phage using Ex-assist  helper phage and propagated in
                      XLOLR cells."
ORIGIN

Alignment Scores:
Pred. No.:                2.01e+03        Length:         355
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       2               Gaps:           0

US-09-858-332G-28 (1-6) x BE240393 (1-355)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         67 CATAACCATAACCATAAC 50

RESULT 148
BI305914/c
LOCUS       BI305914                357 bp    mRNA    linear   EST 20-JUL-2001
DEFINITION  NL_2_G16 Drought stress (leaf) Oryza sativa (indica cultivar-group)
            cDNA clone NL_2_G16 3', mRNA sequence.
ACCESSION   BI305914
VERSION     BI305914.1  GI:14981236
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 357)
  AUTHORS   Reddy,A.R., Ramakrishna,W., Chandrasekhar,A., Nagabhushan,I.,
            Ravindrababu,P. and Bennetzen,J.L.
  TITLE     Novel EST enrichment with normalized cDNA libraries from drought
            stressed rice (Oryza sativa L.cv Nagina 22)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Reddy AR
            Department of Plant Sciences, School of Life Sciences
            University of Hyderabad
            P.O. Central University, Hyderabad-500 046, A.P, India
            Tel: 0091-40-3010265
            Fax: 0091-40-3010145
            Email: arjulsl@uohyd.ernet.in
            Insert Length: 357   Std Error: 0.00
            Plate: 2  row: G  column: 16
            Seq primer: GTAAAACGACGGCCAGTG.
FEATURES              Location/Qualifiers
     source           1. .357
                      /organism="Oryza sativa (indica cultivar-group)"
                      /mol_type="mRNA"
                      /cultivar="Nagina 22"
                      /db_xref="taxon:39946"
                      /clone="NL_2_G16"
                      /tissue_type="Entire leaf tissue"
                      /dev_stage="35 day-old seedlings"
                      /clone_lib="Drought stress (leaf)"
                      /note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized
                      leaf cDNA Library from drought stressed seedlings "
ORIGIN

Alignment Scores:
Pred. No.:                2.02e+03        Length:         357
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       3               Gaps:           0

US-09-858-332G-28 (1-6) x BI305914 (1-357)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        356 CACAACCACAACCACAAC 339

RESULT 149
BI419210
LOCUS       BI419210                357 bp    mRNA    linear   EST 15-AUG-2001
DEFINITION  LjNEST29f8r Lotus japonicus nodule library 5 and 7 week-old Lotus
            corniculatus var. japonicus cDNA 5', mRNA sequence.
ACCESSION   BI419210
VERSION     BI419210.1  GI:15190233
KEYWORDS    EST.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1  (bases 1 to 357)
  AUTHORS   Colebatch,G., Freund,S., Trevaskis,B and Udvardi,M.
  TITLE     Lotus japonicus root nodule ESTs: tools for functional genomics
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Udvardi MK
            Molecular Plant Nutrition
            Max Planck Institute of Molecular Plant Physiology
            Am Muehlenberg 1, 14476 Golm, Germany
            Fax: 49 331 567 8250
            Email: udvardi@mpimp-golm.mpg.de
            Seq primer: T7
            High quality sequence stop: 357.
FEATURES              Location/Qualifiers
     source           1. .357
                      /organism="Lotus corniculatus var. japonicus"
                      /mol_type="mRNA"
                      /cultivar="Gifu (B-129)"
                      /db_xref="taxon:34305"
                      /dev_stage="5 and 7 week-old plants"
                      /clone_lib="Lotus japonicus nodule library 5 and 7
                      week-old"
                      /note="Organ: Nodule; Vector: pSPORT1; Site_1: SalI;
                      Site_2: NotI; The library was prepared using mRNA
                      extracted from nodules of 5 and 7 week-old Lotus plants.
                      Nodules were induced by, and contained Mesorhizobium
                      strain R7A."
ORIGIN

Alignment Scores:
Pred. No.:                2.02e+03        Length:         357
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
```

```
Query Match:            100.00%       Indels:        0
DB:                     3             Gaps:          0

US-09-858-332G-28 (1-6) x BI419210 (1-357)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         43 CATAATCATAACCATAAT 60

RESULT 150
CD852815/c
LOCUS       CD852815                357 bp    mRNA    linear   EST 11-JUL-2003
DEFINITION  DH0ALL9ZH02ZZM1 HaDevR5 Helianthus annuus cDNA clone HaDevR59H02,
            mRNA sequence.
ACCESSION   CD852815
VERSION     CD852815.1  GI:32536631
KEYWORDS    EST.
SOURCE      Helianthus annuus (common sunflower)
  ORGANISM  Helianthus annuus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; campanulids; Asterales; Asteraceae; Asteroideae;
            Heliantheae; Helianthus.
REFERENCE   1  (bases 1 to 357)
  AUTHORS   Genoplante.
  TITLE     Genoplante, a major partnership french program in plant genomics
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Genoplante
            Genoplante
            93, rue Henri Rochefort 91025 EVRY CEDEX France
            Tel: 33 1 69 47 54 00
            Fax: 33 1 69 47 54 10
            This sequence has been generated in the framework of the french
            plant genomics programme 'Genoplante' (http://www.genoplante.com
            and http://genoplante-info.infobiogen.fr).
FEATURES             Location/Qualifiers
     source          1. .357
                     /organism="Helianthus annuus"
                     /mol_type="mRNA"
                     /cultivar="psc8"
                     /db_xref="taxon:4232"
                     /clone="HaDevR59H02"
                     /tissue_type="4 days after self-pollination embryo"
                     /clone_lib="HaDevR5"
ORIGIN

Alignment Scores:
Pred. No.:              2.02e+03      Length:        357
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-28 (1-6) x CD852815 (1-357)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        342 CATAATCATAATCATAAT 325

RESULT 151
W62873
LOCUS       W62873                  357 bp    mRNA    linear   EST 07-JUN-1996
DEFINITION  md88a07.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:385044 5' similar to gb:M38188 OVARIAN GRANULOSA CELL
            13.0 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION   W62873
VERSION     W62873.1  GI:1369622
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 357)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:236876
            Seq primer: ETPrimer
            High quality sequence stop: 351.
FEATURES             Location/Qualifiers
     source          1. .357
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:385044"
                     /sex="unknown"
                     /tissue_type="embryo"
                     /dev_stage="13.5-14.5dpc total fetus"
                     /lab_host="DH10B"
                     /clone_lib="Soares mouse embryo NbME13.5 14.5"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTTT
                     T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                     14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                     State Univ., from 2 ]; double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector. Library went through one round of
                     normalization, and was constructed by Bento Soares and
                     M.Fatima Bonaldo.  "
ORIGIN

Alignment Scores:
Pred. No.:              2.02e+03      Length:        357
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     8             Gaps:          0

US-09-858-332G-28 (1-6) x W62873 (1-357)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        270 CACAACCATAACCACAAC 287

RESULT 152
AU268115
LOCUS       AU268115                358 bp    mRNA    linear   EST 26-APR-2004
DEFINITION  AU268115 VS Dictyostelium discoideum cDNA clone VSH879 5', mRNA
            sequence.
ACCESSION   AU268115
VERSION     AU268115.1  GI:20526913
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 358)
```

```
  AUTHORS   Urushihara,H., Morio,T., Saito,T., Kohara,Y., Koriki,E., Ochiai,H.,
            Maeda,M., Williams,J.G., Takeuchi,I. and Tanaka,Y.
  TITLE     Analyses of cDNAs from growth and slug stages of Dictyostelium
            discoideum
  JOURNAL   Nucleic Acids Res. 32 (5), 1647-1653 (2004)
   PUBMED   15010511
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .358
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="VSH879"
                     /sex="mat A"
                     /dev_stage="vegetative"
                     /clone_lib="VS"
ORIGIN

Alignment Scores:
Pred. No.:                2.03e+03        Length:          358
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       1               Gaps:            0

US-09-858-332G-28 (1-6) x AU268115 (1-358)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        278 CACAATCACAATCACAAC 295

RESULT 153
BM300074
LOCUS       BM300074                358 bp    mRNA    linear   EST 22-JAN-2002
DEFINITION  MCR054H05_24676 Ice plant Lambda Uni-Zap XR expression library, 48
            hours NaCl treatment prescreened for removal of highly abundant
            transcripts Mesembryanthemum crystallinum cDNA clone MCR054H05 5,
            mRNA sequence.
ACCESSION   BM300074
VERSION     BM300074.1  GI:18022449
KEYWORDS    EST.
SOURCE      Mesembryanthemum crystallinum (common iceplant)
  ORGANISM  Mesembryanthemum crystallinum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            Caryophyllales; Aizoaceae; Mesembryanthemum.
REFERENCE   1  (bases 1 to 358)
  AUTHORS   Cushman, J.C.
  TITLE     An expressed sequence tag database for the common ice plant,
            Mesembryanthemum crystallinum
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Cushman JC
            Department of Biochemistry
            University of Nevada
            MS200, Reno, NV 89557-0014, USA
            Tel: 775-784-1918
            Fax: 775-784-1650
            Email: jcushman@unr.edu
            PCR PRimers
            FORWARD: T3 20mer
            BACKWARD: T7 21mer
            Plate: 054  row: H  column: 05
            Seq primer: T3 20mer
            High quality sequence stop: 358.
FEATURES             Location/Qualifiers
     source          1. .358
                     /organism="Mesembryanthemum crystallinum"
                     /mol_type="mRNA"
                     /db_xref="taxon:3544"
                     /clone="MCR054H05"
                     /tissue_type="leaf"
                     /dev_stage="six-week-old"
                     /clone_lib="Ice plant Lambda Uni-Zap XR expression
                     library, 48 hours NaCl treatment prescreened for removal
                     of highly abundant transcripts"
                     /note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI"
ORIGIN

Alignment Scores:
Pred. No.:                2.03e+03        Length:          358
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       3               Gaps:            0

US-09-858-332G-28 (1-6) x BM300074 (1-358)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         16 CACAACCACAACCACAAC 33

RESULT 154
BW840485/c
LOCUS       BW840485                358 bp    mRNA    linear   EST 23-MAY-2005
DEFINITION  BW840485 Amphioxus Branchiostoma floridae unpublished cDNA library,
            neurula whole animal Branchiostoma floridae cDNA clone bbne012a05
            5', mRNA sequence.
ACCESSION   BW840485
VERSION     BW840485.1  GI:66448701
KEYWORDS    EST.
SOURCE      Branchiostoma floridae (Florida lancelet)
  ORGANISM  Branchiostoma floridae
            Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
            Branchiostoma.
REFERENCE   1  (bases 1 to 358)
  AUTHORS   Yu,J., Holland,L.Z., Shin-i,T., Kohara,Y., Satou,Y. and Satoh,N.
  TITLE     Expressed genes in Branchiostoma floridae
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .358
                     /organism="Branchiostoma floridae"
                     /mol_type="mRNA"
                     /db_xref="taxon:7739"
                     /clone="bbne012a05"
                     /tissue_type="whole animal"
                     /dev_stage="neurula"
                     /clone_lib="Amphioxus Branchiostoma floridae unpublished
                     cDNA library, neurula whole animal"
ORIGIN

Alignment Scores:
Pred. No.:                2.03e+03        Length:          358
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       5               Gaps:            0

US-09-858-332G-28 (1-6) x BW840485 (1-358)
```

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        146 CATAATCATAATCATAAT 129

RESULT 155
BJ340157/c
LOCUS       BJ340157                361 bp    mRNA    linear   EST 07-MAR-2002
DEFINITION  BJ340157 Dictyostelium discoideum cDNA library, AF Dictyostelium
            discoideum cDNA clone dda11k23 3', mRNA sequence.
ACCESSION   BJ340157
VERSION     BJ340157.1  GI:19248519
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 361)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the aggregation
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .361
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dda11k23"
                     /sex="mat A"
                     /dev_stage="Aggregation stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN

Alignment Scores:
Pred. No.:              2.05e+03      Length:         361
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     3             Gaps:           0

US-09-858-332G-28 (1-6) x BJ340157 (1-361)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        334 CACAACCACAATCACAAC 317

RESULT 156
AA409731
LOCUS       AA409731                362 bp    mRNA    linear   EST 26-AUG-1998
DEFINITION  EST01462 Mouse 7.5 dpc embryo ectoplacental cone cDNA library Mus
            musculus cDNA clone C0010C03 5', mRNA sequence.
ACCESSION   AA409731
VERSION     AA409731.1  GI:2066377
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 362)
  AUTHORS   Ko,M.S.H., Threat,T.A., Horton,J.H., Wang,X, Cui,Y., Wang,X.,
            Pryor,E., Paris,J., Wells-Smith,J., Fujiwara,H., Yotsumoto,S. and
            Nakashima,H.
  TITLE     Systematic analyses of mouse genes expressed in embryo implantation
            site
  JOURNAL   Unpublished (1997)
COMMENT     Other_ESTs: EST01461
            Contact: Ko MSH
            Center for Molecular Medicine and Genetics
            Wayne State University
            5047 Gullen Mall, Detroit, MI 48202
            Tel: 3135776708
            Fax: 3135776200
            Email: msko@cmb.biosci.wayne.edu
            Human unknown protein from clone pHGR74 mRNA, complete cds.
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .362
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="ATCC (inhost):1363642"
                     /db_xref="taxon:10090"
                     /clone="C0010C03"
                     /sex="unknown"
                     /tissue_type="ectoplacental cone"
                     /dev_stage="embryonic day 7.5 postconception"
                     /lab_host="DH10B"
                     /clone_lib="Mouse 7.5 dpc embryo ectoplacental cone cDNA
                     library"
                     /note="Organ: embryo; Vector: pSPORT1 (Life Technologies);
                     Site_1: SalI; Site_2: NotI; Total RNAs were extracted from
                     ectoplacental cone of 7.5-dpc embryos.  The
                     double-stranded cDNA was synthesized from total RNAs with
                     an Oligo(dT) primer. The library was constructed by Minoru
                     S. H. Ko."
ORIGIN

Alignment Scores:
Pred. No.:              2.05e+03      Length:         362
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     1             Gaps:           0

US-09-858-332G-28 (1-6) x AA409731 (1-362)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        283 CACAACCATAACCACAAC 300

RESULT 157
CF990498
LOCUS       CF990498                362 bp    mRNA    linear   EST 04-FEB-2005
DEFINITION  22837rsicee_9927.y1 Oryza sativa cv. LYP9 tillering whole plant
            cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF990498
VERSION     CF990498.1  GI:58592190
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 362)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
  PUBMED    15685292
COMMENT     Contact: Yan Zhou
```

```
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 362
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .362
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="tillering"
                     /clone_lib="Oryza sativa cv. LYP9 tillering whole plant
                     cDNA library"
ORIGIN

Alignment Scores:
Pred. No.:              2.05e+03        Length:         362
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     7               Gaps:           0

US-09-858-332G-28 (1-6) x CF990498 (1-362)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        140 CACAACCACAACCACAAC 157

RESULT 158
R40227
LOCUS       R40227                  362 bp    mRNA    linear   EST 22-MAY-1995
DEFINITION  yf93f09.s1 Soares infant brain 1NIB Homo sapiens cDNA clone
            IMAGE:30100 3', mRNA sequence.
ACCESSION   R40227
VERSION     R40227.1  GI:820871
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 362)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     On May 5, 1995 this sequence version replaced gi:797843.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1483
            High quality sequence stops: 212 Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1483   Std Error: 0.00
            Seq primer: Promega -21m13
            High quality sequence stop: 212.
FEATURES             Location/Qualifiers
     source          1. .362
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="GDB:402447"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30100"
                     /sex="female"
                     /dev_stage="73 days post natal"
                     /lab_host="DH10B (ampicillin resistant)"
                     /clone_lib="Soares infant brain 1NIB"
                     /note="Organ: whole brain; Vector: Lafmid BA; Site_1: Not
                     I; Site_2: Hind III; 1st strand cDNA was primed with a Not
                     I - oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTT 3']
                     double-stranded cDNA was ligated to Hind III adaptors
                     (Pharmacia), digested with Not I and directionally cloned
                     into the Not I and Hind III sites of the Lafmid BA vector.
                     Library went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
ORIGIN

Alignment Scores:
Pred. No.:              2.05e+03        Length:         362
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     8               Gaps:           0

US-09-858-332G-28 (1-6) x R40227 (1-362)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        202 CATAATCATAATCATAAT 219

RESULT 159
BE484332
LOCUS       BE484332                363 bp    mRNA    linear   EST 27-MAR-2003
DEFINITION  170900 BARC 5BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BE484332
VERSION     BE484332.1  GI:9603865
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 363)
  AUTHORS   Sonstegard,T., Capuco,A.V., White,J., Van Tassell,C.P.,
            Connor,E.E., Cho,J., Sultana,R., Shade,L., Wray,J.E., Wells,K.D.
            and Quackenbush,J.
  TITLE     Analysis of bovine mammary gland EST and functional annotation of
            the Bos taurus gene index
  JOURNAL   Mamm. Genome 13 (7), 373-379 (2002)
   PUBMED   12140684
COMMENT     Contact: Sonstegard TS
            USDA, ARS, Beltsville Agricultural Research Center
            Bdlg. 200 Rm 2A, Beltsville, MD 20705, USA
            Tel: 301 504 8416
            Fax: 301 504 8414
            Email: tads@lpsi.barc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 25  row: G  column: 10
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .363
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /tissue_type="pooled"
```

```
                     /lab_host="DH10B"
                     /clone_lib="BARC 5BOV"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled mRNA isolated from mammary
                     tissues at eight physiological, developmental, and disease
                     states."
ORIGIN

Alignment Scores:
Pred. No.:              2.06e+03        Length:         363
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-858-332G-28 (1-6) x BE484332 (1-363)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        305 CACAACCACAACCACAAC 322

RESULT 160
BI305814/c
LOCUS       BI305814                363 bp    mRNA    linear   EST 20-JUL-2001
DEFINITION  NL_1_N10 Drought stress (leaf) Oryza sativa (indica cultivar-group)
            cDNA clone NL_1_N10 3', mRNA sequence.
ACCESSION   BI305814
VERSION     BI305814.1  GI:14981123
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 363)
  AUTHORS   Reddy,A.R., Ramakrishna,W., Chandrasekhar,A., Nagabhushan,I.,
            Ravindrababu,P. and Bennetzen,J.L.
  TITLE     Novel EST enrichment with normalized cDNA libraries from drought
            stressed rice (Oryza sativa L.cv Nagina 22)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Reddy AR
            Department of Plant Sciences, School of Life Sciences
            University of Hyderabad
            P.O. Central University, Hyderabad-500 046, A.P, India
            Tel: 0091-40-3010265
            Fax: 0091-40-3010145
            Email: arjulsl@uohyd.ernet.in
            Insert Length: 363   Std Error: 0.00
            Plate: 1  row: N  column: 10
            Seq primer: GTAAAACGACGGCCAGTG.
FEATURES             Location/Qualifiers
     source          1. .363
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nagina 22"
                     /db_xref="taxon:39946"
                     /clone="NL_1_N10"
                     /tissue_type="Entire leaf tissue"
                     /dev_stage="35 day-old seedlings"
                     /clone_lib="Drought stress (leaf)"
                     /note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized
                     leaf cDNA Library from drought stressed seedlings "
ORIGIN

Alignment Scores:
Pred. No.:              2.06e+03        Length:         363
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-858-332G-28 (1-6) x BI305814 (1-363)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        360 CACAACCACAACCACAAC 343

RESULT 161
BH142412/c
LOCUS       BH142412                363 bp    DNA     linear   GSS 16-AUG-2001
DEFINITION  TDGDO50TH cTOG Lycopersicon esculentum genomic clone cTOG24I3,
            genomic survey sequence.
ACCESSION   BH142412
VERSION     BH142412.1  GI:15194980
KEYWORDS    GSS.
SOURCE      Lycopersicon esculentum (Solanum lycopersicum)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 363)
  AUTHORS   van der Hoeven,R., Sun,H., Cho,J., Utterback,T., Ronning,C. and
            Tanksley,S.
  TITLE     Tomato Demethylated Genomic DNA Sequences
  JOURNAL   Unpublished (2001)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html
            tomato demethylated genomic DNA
            Insert Length: 1270   Std Error: 0.00
            Seq primer: M13F-R
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .363
                     /organism="Lycopersicon esculentum"
                     /mol_type="genomic DNA"
                     /cultivar="E6203"
                     /db_xref="taxon:4081"
                     /clone="cTOG24I3"
                     /tissue_type="young leaves"
                     /dev_stage="12-14 weeks post harvest"
                     /lab_host="E.coli JM109"
                     /clone_lib="cTOG"
                     /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                     XhoI; This library was made from short EcoRI digested
                     fragments of the genome of Lycopersicon esculentum ligated
                     into pBS (SK-). The fragments were  cloned into the
                     methylation restrictive E.coli strain JM109 with the
                     purpose of enriching the library for non-methylated DNA
                     fragments.  This procedure may enrich the pool of cloned
                     fragments in JM109 cells for sequences representing
                     expressed genes. Average insert size 1.27 kb."
ORIGIN

Alignment Scores:
Pred. No.:              2.06e+03        Length:         363
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     9               Gaps:           0

US-09-858-332G-28 (1-6) x BH142412 (1-363)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        214 CATAATCATAATCATAAT 197

RESULT 162
BI306083/c
LOCUS       BI306083                364 bp    mRNA    linear   EST 20-JUL-2001
DEFINITION  NL_3_E04 Drought stress (leaf) Oryza sativa (indica cultivar-group)
```

```
            cDNA clone NL_3_E04 3', mRNA sequence.
ACCESSION   BI306083
VERSION     BI306083.1  GI:14981405
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 364)
  AUTHORS   Reddy,A.R., Ramakrishna,W., Chandrasekhar,A., Nagabhushan,I.,
            Ravindrababu,P. and Bennetzen,J.L.
  TITLE     Novel EST enrichment with normalized cDNA libraries from drought
            stressed rice (Oryza sativa L.cv Nagina 22)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Reddy AR
            Department of Plant Sciences, School of Life Sciences
            University of Hyderabad
            P.O. Central University, Hyderabad-500 046, A.P, India
            Tel: 0091-40-3010265
            Fax: 0091-40-3010145
            Email: arjulsl@uohyd.ernet.in
            Insert Length: 364   Std Error: 0.00
            Plate: 3  row: E  column: 04
            Seq primer: GTAAAACGACGGCCAGTG.
FEATURES              Location/Qualifiers
     source           1. .364
                      /organism="Oryza sativa (indica cultivar-group)"
                      /mol_type="mRNA"
                      /cultivar="Nagina 22"
                      /db_xref="taxon:39946"
                      /clone="NL_3_E04"
                      /tissue_type="Entire leaf tissue"
                      /dev_stage="35 day-old seedlings"
                      /clone_lib="Drought stress (leaf)"
                      /note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized
                      leaf cDNA Library from drought stressed seedlings "
ORIGIN

Alignment Scores:
Pred. No.:               2.06e+03       Length:         364
Score:                   42.00          Matches:        6
Percent Similarity:      100.00%        Conservative:   0
Best Local Similarity:   100.00%        Mismatches:     0
Query Match:             100.00%        Indels:         0
DB:                      3              Gaps:           0

US-09-858-332G-28 (1-6) x BI306083 (1-364)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        361 CACAACCACAACCACAAC 344

RESULT 163
BI306635/c
LOCUS       BI306635                364 bp    mRNA    linear   EST 20-JUL-2001
DEFINITION  NL_5_I10 Drought stress (leaf) Oryza sativa (indica cultivar-group)
            cDNA clone NL_5_I10 3', mRNA sequence.
ACCESSION   BI306635
VERSION     BI306635.1  GI:14981957
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 364)
  AUTHORS   Reddy,A.R., Ramakrishna,W., Chandrasekhar,A., Nagabhushan,I.,
            Ravindrababu,P. and Bennetzen,J.L.
  TITLE     Novel EST enrichment with normalized cDNA libraries from drought
            stressed rice (Oryza sativa L.cv Nagina 22)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Reddy AR
            Department of Plant Sciences, School of Life Sciences
            University of Hyderabad
            P.O. Central University, Hyderabad-500 046, A.P, India
            Tel: 0091-40-3010265
            Fax: 0091-40-3010145
            Email: arjulsl@uohyd.ernet.in
            Insert Length: 364   Std Error: 0.00
            Plate: 5  row: I  column: 10
            Seq primer: GTAAAACGACGGCCAGTG.
FEATURES              Location/Qualifiers
     source           1. .364
                      /organism="Oryza sativa (indica cultivar-group)"
                      /mol_type="mRNA"
                      /cultivar="Nagina 22"
                      /db_xref="taxon:39946"
                      /clone="NL_5_I10"
                      /tissue_type="Entire leaf tissue"
                      /dev_stage="35 day-old seedlings"
                      /clone_lib="Drought stress (leaf)"
                      /note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized
                      leaf cDNA Library from drought stressed seedlings "
ORIGIN

Alignment Scores:
Pred. No.:               2.06e+03       Length:         364
Score:                   42.00          Matches:        6
Percent Similarity:      100.00%        Conservative:   0
Best Local Similarity:   100.00%        Mismatches:     0
Query Match:             100.00%        Indels:         0
DB:                      3              Gaps:           0

US-09-858-332G-28 (1-6) x BI306635 (1-364)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        354 CACAACCACAACCACAAC 337

RESULT 164
BI306516/c
LOCUS       BI306516                365 bp    mRNA    linear   EST 20-JUL-2001
DEFINITION  NL_5_A01 Drought stress (leaf) Oryza sativa (indica cultivar-group)
            cDNA clone NL_5_A01 3', mRNA sequence.
ACCESSION   BI306516
VERSION     BI306516.1  GI:14981838
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 365)
  AUTHORS   Reddy,A.R., Ramakrishna,W., Chandrasekhar,A., Nagabhushan,I.,
            Ravindrababu,P. and Bennetzen,J.L.
  TITLE     Novel EST enrichment with normalized cDNA libraries from drought
            stressed rice (Oryza sativa L.cv Nagina 22)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Reddy AR
            Department of Plant Sciences, School of Life Sciences
            University of Hyderabad
            P.O. Central University, Hyderabad-500 046, A.P, India
            Tel: 0091-40-3010265
            Fax: 0091-40-3010145
            Email: arjulsl@uohyd.ernet.in
            Insert Length: 365   Std Error: 0.00
            Plate: 5  row: A  column: 01
            Seq primer: GTAAAACGACGGCCAGTG.
FEATURES              Location/Qualifiers
     source           1. .365
                      /organism="Oryza sativa (indica cultivar-group)"
                      /mol_type="mRNA"
                      /cultivar="Nagina 22"
                      /db_xref="taxon:39946"
                      /clone="NL_5_A01"
```

```
                     /tissue_type="Entire leaf tissue"
                     /dev_stage="35 day-old seedlings"
                     /clone_lib="Drought stress (leaf)"
                     /note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized
                     leaf cDNA Library from drought stressed seedlings "
ORIGIN

Alignment Scores:
Pred. No.:              2.07e+03        Length:         365
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-858-332G-28 (1-6) x BI306516 (1-365)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        364 CACAACCACAACCACAAC 347

RESULT 165
AI157489
LOCUS       AI157489                366 bp    mRNA    linear   EST 30-SEP-1998
DEFINITION  ue56e05.r1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
            IMAGE:1495136 5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0
            KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION   AI157489
VERSION     AI157489.1  GI:3685958
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 366)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:932740
            Putative full length read
            vector to vector length is 614
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 361.
FEATURES             Location/Qualifiers
     source          1. .366
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1495136"
                     /sex="female (lactating)"
                     /tissue_type="mammary gland"
                     /lab_host="DH10B"
                     /clone_lib="Soares_mammary_gland_NMLMG"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from mammary
                     gland tissue from a lactating female, and was then primed
                     with a Not I - oligo(dT) primer. Double-stranded  cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I  and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector.  Library is normalized.  Library
                     was constructed by Bento Soares and M. Fatima Bonaldo. "
ORIGIN

Alignment Scores:
Pred. No.:              2.08e+03        Length:         366
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AI157489 (1-366)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         90 CACAACCATAACCACAAC 107

RESULT 166
BP595103/c
LOCUS       BP595103                366 bp    mRNA    linear   EST 23-JUN-2004
DEFINITION  BP595103 RAFL15 Arabidopsis thaliana cDNA clone RAFL15-28-I08 3',
            mRNA sequence.
ACCESSION   BP595103
VERSION     BP595103.1  GI:49162571
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 366)
  AUTHORS   Seki,M., Narusaka,M., Kamiya,A., Ishida,J., Satou,M., Sakurai,T.,
            Nakajima,M., Enju,A., Akiyama,K., Oono,Y., Muramatsu,M.,
            Hayashizaki,Y., Kawai,J., Carninci,P., Itoh,M., Ishii,Y.,
            Arakawa,T., Shibata,K., Shinagawa,A. and Shinozaki,K.
  TITLE     Functional annotation of a full-length Arabidopsis cDNA collection
  JOURNAL   Science 296 (5565), 141-145 (2002)
   PUBMED   11910074
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            reversed clone; Please visit our web site
            (http://pfgweb.gsc.riken.go.jp/) for further details.
FEATURES             Location/Qualifiers
     source          1. .366
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="RAFL15-28-I08"
                     /tissue_type="mixture of silique and flower"
                     /lab_host="DH10B"
                     /clone_lib="RAFL15"
                     /note="Site_1: BamHI; Site_2: SalI"
ORIGIN

Alignment Scores:
Pred. No.:              2.08e+03        Length:         366
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-858-332G-28 (1-6) x BP595103 (1-366)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        306 CACAACCACAACCACAAC 289
```

```
RESULT 167
BY062997
LOCUS       BY062997                366 bp    mRNA    linear   EST 06-DEC-2002
DEFINITION  BY062997 RIKEN full-length enriched, 17 days embryo heart Mus
            musculus cDNA clone I920025F19 5', mRNA sequence.
ACCESSION   BY062997
VERSION     BY062997.1  GI:26167750
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 366)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
            Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
            Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
            Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
            Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
            Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
            Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
            Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
            Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
            Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
            Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
            Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
            Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
            Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
            Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
            Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
            Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
            Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
            Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
            Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
            Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
            Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
            Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
            Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
            Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
            Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
   PUBMED   12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
            Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
            Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,
            Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,
            Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and
            Hayashizaki,Y. Direct Submission
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
             cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .366
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="I920025F19"
                     /tissue_type="heart"
                     /dev_stage="17 days embryo"
                     /clone_lib="RIKEN full-length enriched, 17 days embryo
                     heart"
ORIGIN

Alignment Scores:
Pred. No.:              2.08e+03        Length:         366
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     5               Gaps:           0

US-09-858-332G-28 (1-6) x BY062997 (1-366)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        301 CACAACCATAACCACAAC 318

RESULT 168
CN926563
LOCUS       CN926563                366 bp    mRNA    linear   EST 07-JUN-2004
DEFINITION  000501AEPA001918HT (AEPA) Pinkie expanding leaf Malus x domestica
            cDNA clone AEPA001918, mRNA sequence.
ACCESSION   CN926563
VERSION     CN926563.1  GI:48399376
KEYWORDS    EST.
SOURCE      Malus x domestica
  ORGANISM  Malus x domestica
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Rosales; Rosaceae; Maloideae; Malus.
REFERENCE   1  (bases 1 to 366)
  AUTHORS   Beuning,L., Bowen,J., Crowhurst,R., Gleave,A., Janssen,B.,
            McArtney,S., Newcomb,R., Ross,G., Snowden,K., Walton,E. and Yauk,Y.
  TITLE     HortResearch Apple EST Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Gleave,A.
            Sequencing Facility
            The Horticulture and Food Research Institute of New Zealand Ltd
            120 Mt Albert Rd, Mt Albert, Auckland, New Zealand
            Tel: 00 64 09 815 4200
            Fax: 00 64 09 815 4201
            Email: est@hortresearch.co.nz.
FEATURES             Location/Qualifiers
     source          1. .366
                     /organism="Malus x domestica"
                     /mol_type="mRNA"
                     /db_xref="taxon:3750"
                     /clone="AEPA001918"
                     /tissue_type="Leaf"
                     /dev_stage="Expanding"
                     /clone_lib="(AEPA) Pinkie expanding leaf"
                     /note="Vector: pBK-CMV; Library sequenced by Genesis
                     Research & Development"
ORIGIN

Alignment Scores:
Pred. No.:              2.08e+03        Length:         366
Score:                  42.00           Matches:        6
```

```
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     7              Gaps:          0

US-09-858-332G-28 (1-6) x CN926563 (1-366)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        185 CACAACCACAACCACAAC 202

RESULT 169
R41647
LOCUS       R41647                   366 bp    mRNA    linear   EST 22-MAY-1995
DEFINITION  yf88a10.s1 Soares infant brain 1NIB Homo sapiens cDNA clone
            IMAGE:29528 3', mRNA sequence.
ACCESSION   R41647
VERSION     R41647.1  GI:816942
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 366)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     On May 8, 1995 this sequence version replaced gi:799871.
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 354
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: Promega -21m13
            High quality sequence stop: 354.
FEATURES             Location/Qualifiers
     source          1. .366
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="GDB:401875"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:29528"
                     /sex="female"
                     /dev_stage="73 days post natal"
                     /lab_host="DH10B (ampicillin resistant)"
                     /clone_lib="Soares infant brain 1NIB"
                     /note="Organ: whole brain; Vector: Lafmid BA; Site_1: Not
                     I; Site_2: Hind III; 1st strand cDNA was primed with a Not
                     I - oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTTT 3']
                     double-stranded cDNA was ligated to Hind III adaptors
                     (Pharmacia), digested with Not I and directionally cloned
                     into the Not I and Hind III sites of the Lafmid BA vector.
                     Library went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
ORIGIN

Alignment Scores:
Pred. No.:              2.08e+03       Length:        366
Score:                  42.00          Matches:       6
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     8              Gaps:          0

US-09-858-332G-28 (1-6) x R41647 (1-366)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        199 CATAATCATAATCATAAT 216

RESULT 170
BY301184
LOCUS       BY301184                 367 bp    mRNA    linear   EST 11-DEC-2002
DEFINITION  BY301184 RIKEN full-length enriched, 14.5 days embryo RP+/+
            Rathke's pouches Mus musculus cDNA clone K720013L20 5', mRNA
            sequence.
ACCESSION   BY301184
VERSION     BY301184.1  GI:26491521
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 367)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
            Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
            Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
            Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
            Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
            Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
            Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
            Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
            Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
            Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
            Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
            Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
            Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
            Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
            Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
            Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
            Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
            Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
            Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
            Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
            Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
            Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
            Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
            Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
            Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
            Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
   PUBMED   12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
            Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
            Miyazaki,A., Murata,M., Nakamura,K., Numazaki,R.,
            Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,
            Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and
            Hayashizaki,Y. Direct Submission
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
```

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Tissues were provided by Michelle Brinkmeier and Sally Camper (Dept. Human Genetics University of Michigan Medical School 4301 MSRB 3 1500 W. Medical Center Dr. Ann Arbor,MI 48109-0638 USA) whose assistance we gratefully acknowledge.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

FEATURES Location/Qualifiers
source 1. .367
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="K720013L20"
/tissue_type="Rathke's pouches"
/dev_stage="14.5 days embryo RP+/+"
/clone_lib="RIKEN full-length enriched, 14.5 days embryo RP+/+ Rathke's pouches"

ORIGIN

Alignment Scores:

Pred. No.:	2.08e+03	Length:	367
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	5	Gaps:	0

US-09-858-332G-28 (1-6) x BY301184 (1-367)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        284 CACAACCATAACCACAAC 301
```

RESULT 171
AJ761032/c
LOCUS AJ761032 368 bp mRNA linear EST 03-JUN-2005
DEFINITION AJ761032 G00005 Gerbera hybrid cv. 'Terra Regina' cDNA clone G0000500015D09F1, mRNA sequence.
ACCESSION AJ761032
VERSION AJ761032.1 GI:62942087
KEYWORDS EST.
SOURCE Gerbera hybrid cv. 'Terra Regina'
ORGANISM Gerbera hybrid cv. 'Terra Regina'
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; campanulids; Asterales; Asteraceae; Mutisioideae; Mutisieae; Gerbera.
REFERENCE 1 (bases 1 to 368)
AUTHORS Laitinen,R.A., Immanen,J., Auvinen,P., Rudd,S., Alatalo,E.R., Paulin,L., Ainasoja,M., Kotilainen,M., Koskela,S., Teeri,T.H. and Elomaa,P.
TITLE Analysis of the floral transcriptome uncovers new regulators of organ determination and gene families related to flower organ differentiation in Gerbera hybrida (Asteraceae)
JOURNAL Genome Res. 15 (4), 475-486 (2005)
PUBMED 15781570
COMMENT Contact: Alatalo ER
Institute of Biotechnology
Viikinkaari 9, P.O. Box 56, 00014 University of Helsinki, FINLAND.

FEATURES Location/Qualifiers
source 1. .368
/organism="Gerbera hybrid cv. 'Terra Regina'"
/mol_type="mRNA"
/cultivar="Terra Regina"
/db_xref="taxon:226891"
/clone="G0000500015D09F1"
/tissue_type="leaf, transgenic line 35S-gaga2"
/clone_lib="G00005"

ORIGIN

Alignment Scores:

Pred. No.:	2.09e+03	Length:	368
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0

US-09-858-332G-28 (1-6) x AJ761032 (1-368)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        310 CATAATCATAATCATAAT 293
```

RESULT 172
BY083645
LOCUS BY083645 368 bp mRNA linear EST 07-DEC-2002
DEFINITION BY083645 RIKEN full-length enriched, 10 days neonate thymus Mus musculus cDNA clone K630044K19 5', mRNA sequence.
ACCESSION BY083645
VERSION BY083645.1 GI:26203392
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 368)
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
PUBMED 12466851
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216

```
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
            Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
            Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,
            Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,
            Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and
            Hayashizaki,Y. Direct Submission
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
             cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .368
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="K630044K19"
                     /tissue_type="thymus"
                     /dev_stage="10 days neonate"
                     /clone_lib="RIKEN full-length enriched, 10 days neonate
                     thymus"
ORIGIN

Alignment Scores:
Pred. No.:              2.09e+03       Length:         368
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     5              Gaps:           0

US-09-858-332G-28 (1-6) x BY083645 (1-368)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        295 CACAACCATAACCACAAC 312

RESULT 173
CF963928
LOCUS       CF963928                368 bp    mRNA    linear   EST 15-JAN-2004
DEFINITION  13692rsiced_4143.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF963928
VERSION     CF963928.1  GI:38479839
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 368)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 368
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .368
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="trefoil"
                     /clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA
                     library"
ORIGIN

Alignment Scores:
Pred. No.:              2.09e+03       Length:         368
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     7              Gaps:           0

US-09-858-332G-28 (1-6) x CF963928 (1-368)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        218 CACAACCACAACCACAAC 235

RESULT 174
AU039300
LOCUS       AU039300                369 bp    mRNA    linear   EST 29-MAR-1999
DEFINITION  AU039300 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
            discoideum cDNA clone SLH502, mRNA sequence.
ACCESSION   AU039300
VERSION     AU039300.1  GI:4008281
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 369)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     The Dictyostelium developmental cDNA project: generation and
            analysis of expressed sequence tags from the first-finger stage of
            development
  JOURNAL   DNA Res. 5 (6), 335-340 (1998)
   PUBMED   10048482
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp
            PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1. .369
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
```

```
                     /db_xref="taxon:44689"
                     /clone="SLH502"
                     /dev_stage="slug"
                     /clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
ORIGIN

Alignment Scores:
Pred. No.:                2.09e+03        Length:         369
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       1               Gaps:           0

US-09-858-332G-28 (1-6) x AU039300 (1-369)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        138 CATAATCATAATCATAAT 155

RESULT 175
BJ356615/c
LOCUS       BJ356615                369 bp    mRNA    linear   EST 07-MAR-2002
DEFINITION  BJ356615 Dictyostelium discoideum cDNA library, AF Dictyostelium
            discoideum cDNA clone dda60j24 3', mRNA sequence.
ACCESSION   BJ356615
VERSION     BJ356615.1  GI:19256210
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 369)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the aggregation
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .369
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dda60j24"
                     /sex="mat A"
                     /dev_stage="Aggregation stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN

Alignment Scores:
Pred. No.:                2.09e+03        Length:         369
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       3               Gaps:           0

US-09-858-332G-28 (1-6) x BJ356615 (1-369)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        312 CACAATCACAACCACAAC 295

RESULT 176
BP630263/c
LOCUS       BP630263                369 bp    mRNA    linear   EST 27-JUN-2004
DEFINITION  BP630263 RAFL17 Arabidopsis thaliana cDNA clone RAFL17-19-H13 3',
            mRNA sequence.
ACCESSION   BP630263
VERSION     BP630263.1  GI:49281515
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 369)
  AUTHORS   Seki,M., Narusaka,M., Kamiya,A., Ishida,J., Satou,M., Sakurai,T.,
            Nakajima,M., Enju,A., Akiyama,K., Oono,Y., Muramatsu,M.,
            Hayashizaki,Y., Kawai,J., Carninci,P., Itoh,M., Ishii,Y.,
            Arakawa,T., Shibata,K., Shinagawa,A. and Shinozaki,K.
  TITLE     Functional annotation of a full-length Arabidopsis cDNA collection
  JOURNAL   Science 296 (5565), 141-145 (2002)
  PUBMED    11910074
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            reversed clone; Please visit our web site
            (http://pfgweb.gsc.riken.go.jp/) for further details.
FEATURES             Location/Qualifiers
     source          1. .369
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="RAFL17-19-H13"
                     /lab_host="DH10B"
                     /clone_lib="RAFL17"
                     /note="Site_1: BamHI; Site_2: SalI; Subtraction Library.
                     The sequence was obtained from samples subjected to
                     dehydration-treated (1, 2, 5, 10 and 24 hr) and
                     rehydration-treated (1, 2, 5, 10, and 24 hr after
                     dehydration treatment)"
ORIGIN

Alignment Scores:
Pred. No.:                2.09e+03        Length:         369
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       3               Gaps:           0

US-09-858-332G-28 (1-6) x BP630263 (1-369)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        297 CACAACCACAACCACAAC 280

RESULT 177
BF292801
LOCUS       BF292801                370 bp    mRNA    linear   EST 17-NOV-2000
DEFINITION  WHE2165_A09_B17ZS Triticum turgidum L. var. durum (durum wheat)
            whole plant cDNA library Triticum turgidum cDNA clone
            WHE2165_A09_B17, mRNA sequence.
ACCESSION   BF292801
VERSION     BF292801.1  GI:11223865
KEYWORDS    EST.
SOURCE      Triticum turgidum
  ORGANISM  Triticum turgidum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 370)
  AUTHORS   Akhunov,E., Anderson,O.D., Chao,S., Chin,A., Choi,D.W., Close,T.J.,
            Fenton,R.D., Han,P.S., Hsia,C.C., Kang,Y., Kianian,P., Lazo,G.R.,
            Miller,R., Otto,C., Rausch,C.J., Seaton,C.L., Simons,K., Tong,J.C.
```

```
          and Zhang,D.
  TITLE   The structure and function of the expressed portion of the wheat
          genomes - Whole plant cDNA library from Triticum turgidum L. var.
          durum
  JOURNAL Unpublished (2000)
COMMENT   Contact: Olin Anderson
          US Department of Agriculture, Agriculture Research Service, Pacific
          West Area, Western Regional Research Center
          800 Buchanan Street, Albany, CA 94710, USA
          Tel: 5105595773
          Fax: 5105595818
          Email: oandersn@pw.usda.gov
          Sequence have been trimmed to remove vector sequence and low
          quality sequence with phred score less than 20
          Seq primer: Stratagene SK primer.
FEATURES             Location/Qualifiers
     source          1. .370
                     /organism="Triticum turgidum"
                     /mol_type="mRNA"
                     /cultivar="Langdon-16"
                     /db_xref="taxon:4571"
                     /clone="WHE2165_A09_B17"
                     /tissue_type="All tissues"
                     /dev_stage="Different growth stages"
                     /lab_host="E. coli SOLR"
                     /clone_lib="Triticum turgidum L. var. durum (durum wheat)
                     whole plant cDNA library"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; Plants were grown in a growth
                     chamber at North Dakota State University (Kianian, Otto,
                     Simons). Tissues collected from seven-day etiolated
                     seedling leaf, stem, root and seed; leaf from plant at
                     fourth leaf stage; spike from pre-anthesis through 20
                     days after anthesis; flag leaf; leaf and stem tissue from
                     tillers. and root. Total RNA and poly(A) RNA were
                     prepared from each tissue and then pooled, a cDNA library
                     was made, and the cDNA clones were in vivo excised to give
                     pBluescript phagemids in the TJ Close lab (Akhunov, Chin,
                     Choi, Close, Fenton, Kianian, Otto, Simons, Zhang) at the
                     University of California, Riverside. Plasmid DNA
                     preparations and DNA sequencing were performed in the OD
                     Anderson lab (all other authors)."
ORIGIN

Alignment Scores:
Pred. No.:              2.1e+03       Length:         370
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     2             Gaps:           0

US-09-858-332G-28 (1-6) x BF292801 (1-370)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        308 CACAACCACAACCACAAT 325

RESULT 178
C93099
LOCUS       C93099                   370 bp    mRNA    linear   EST 12-JUL-1999
DEFINITION  C93099 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
            discoideum cDNA clone SSM554, mRNA sequence.
ACCESSION   C93099
VERSION     C93099.1  GI:3074975
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 370)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     The Dictyostelium developmental cDNA project: generation and
            analysis of expressed sequence tags from the first-finger stage of
            development
  JOURNAL   DNA Res. 5 (6), 335-340 (1998)
   PUBMED   10048482
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp
            PROJECT = 'Dictyostelium discoideum cDNA project in Japan'
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .370
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SSM554"
                     /dev_stage="slug"
                     /clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
ORIGIN

Alignment Scores:
Pred. No.:              2.1e+03       Length:         370
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     5             Gaps:           0

US-09-858-332G-28 (1-6) x C93099 (1-370)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        138 CATAATCATAATCATAAT 155

RESULT 179
BF643583
LOCUS       BF643583                 371 bp    mRNA    linear   EST 20-DEC-2000
DEFINITION  NF092A09EC1F1069 Elicited cell culture Medicago truncatula cDNA
            clone NF092A09EC 5', mRNA sequence.
ACCESSION   BF643583
VERSION     BF643583.1  GI:11908804
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 371)
  AUTHORS   Torres-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,
            Flores,H.R., Inman,J.T., Weller,J.W. and May,G.D.
  TITLE     Expressed Sequence Tags from the Samuel Roberts Noble Foundation -
            Center for Medicago Genomics Research
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Dixon RA
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA
            Tel: 580 221 7302
            Fax: 580 221 7380
            Email: radixon@noble.org
            Insert Length: 371   Std Error: 0.00
            Plate: 092  row: A  column: 09
            Seq primer: TCACACAGGAAACAGCTATGAC.
FEATURES             Location/Qualifiers
     source          1. .371
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
```

```
                     /db_xref="taxon:3880"
                     /clone="NF092A09EC"
                     /tissue_type="Cell cultures derived from root tissues"
                     /dev_stage="Cell suspensions were subcultured every 14
                     days. Cells were induced six days after subculture"
                     /clone_lib="Elicited cell culture"
                     /note="Vector: Lambda Zap; Cells were induced with yeast
                     cell wall extracts equivalent to  50ug/ml glucose in the
                     final concentration. Samples were taken at  0.5, 1, 12 and
                     24 hours after induction. Equal amounts of RNA from  each
                     time point were pooled and used for mRNA isolation."
ORIGIN

Alignment Scores:
Pred. No.:                2.11e+03        Length:         371
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       2               Gaps:           0

US-09-858-332G-28 (1-6) x BF643583 (1-371)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        160 CATAATCACAATCACAAT 177

RESULT 180
BE521566
LOCUS       BE521566                  371 bp    mRNA    linear   EST 19-MAR-2001
DEFINITION  M20D1STM Arabidopsis developing seed Arabidopsis thaliana cDNA
            clone M20D1 5', mRNA sequence.
ACCESSION   BE521566
VERSION     BE521566.1  GI:9779544
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 371)
  AUTHORS   White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de
            Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.
  TITLE     A new set of Arabidopsis expressed sequence tags from developing
            seeds. The metabolic pathway from carbohydrates to seed oil
  JOURNAL   Plant Physiol. 124 (4), 1582-1594 (2000)
   PUBMED   11115876
COMMENT     Contact: Benning, C
            Dept. of Biochemistry & Molecular Biology
            Michigan State University
            224 Biochemistry, Michigan State University, East Lansing, MI
            48824, USA
            Tel: 517 355 1609
            Fax: 517 353 9334
            Email: benning@msu.edu
            Michigan State University DNA Sequencing Facility Arabidopsis
            Biological Resource Center, The Ohio State University, 309 Botany &
            Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:
            6142920603 TEL: 6142929371.
FEATURES             Location/Qualifiers
     source          1. .371
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /ecotype="Columbia"
                     /db_xref="taxon:3702"
                     /clone="M20D1"
                     /tissue_type="seed"
                     /dev_stage="5-13 days after flowering"
                     /lab_host="E.coli"
                     /clone_lib="Arabidopsis developing seed"
                     /note="Organ: Developing seed; Vector: pBluescript SK-;
                     Site_1: EcoRI; Site_2: XhoII"
ORIGIN

Alignment Scores:
Pred. No.:                2.11e+03        Length:         371
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       2               Gaps:           0

US-09-858-332G-28 (1-6) x BE521566 (1-371)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        264 CACAACCACAACCACAAT 281

RESULT 181
BP623064
LOCUS       BP623064                  371 bp    mRNA    linear   EST 26-JUN-2004
DEFINITION  BP623064 RAFL17 Arabidopsis thaliana cDNA clone RAFL17-05-G22 3',
            mRNA sequence.
ACCESSION   BP623064
VERSION     BP623064.1  GI:49274246
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 371)
  AUTHORS   Seki,M., Narusaka,M., Kamiya,A., Ishida,J., Satou,M., Sakurai,T.,
            Nakajima,M., Enju,A., Akiyama,K., Oono,Y., Muramatsu,M.,
            Hayashizaki,Y., Kawai,J., Carninci,P., Itoh,M., Ishii,Y.,
            Arakawa,T., Shibata,K., Shinagawa,A. and Shinozaki,K.
  TITLE     Functional annotation of a full-length Arabidopsis cDNA collection
  JOURNAL   Science 296 (5565), 141-145 (2002)
   PUBMED   11910074
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            reversed clone; Please visit our web site
            (http://pfgweb.gsc.riken.go.jp/) for further details.
FEATURES             Location/Qualifiers
     source          1. .371
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="RAFL17-05-G22"
                     /lab_host="DH10B"
                     /clone_lib="RAFL17"
                     /note="Site_1: BamHI; Site_2: SalI; Subtraction Library.
                     The sequence was obtained from samples subjected to
                     dehydration-treated (1, 2, 5, 10 and 24 hr) and
                     rehydration-treated (1, 2, 5, 10, and 24 hr after
                     dehydration treatment)"
ORIGIN

Alignment Scores:
Pred. No.:                2.11e+03        Length:         371
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       3               Gaps:           0

US-09-858-332G-28 (1-6) x BP623064 (1-371)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         61 CACAATCATAATCATAAT 78
```

```
RESULT 182
CF991958
LOCUS       CF991958                372 bp    mRNA    linear   EST 04-FEB-2005
DEFINITION  17088rsicee_14776.y1 Oryza sativa cv. LYP9 tillering whole plant
            cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF991958
VERSION     CF991958.1  GI:58593650
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 372)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 372
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .372
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="tillering"
                     /clone_lib="Oryza sativa cv. LYP9 tillering whole plant
                     cDNA library"
ORIGIN

Alignment Scores:
Pred. No.:              2.11e+03        Length:         372
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     7               Gaps:           0

US-09-858-332G-28 (1-6) x CF991958 (1-372)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        205 CACAACCACAACCACAAC 222

RESULT 183
AI016228/c
LOCUS       AI016228                373 bp    mRNA    linear   EST 27-AUG-1998
DEFINITION  ot83h12.s1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone
            IMAGE:1623431 3' similar to gb:X07868_rna1 PUTATIVE INSULIN-LIKE
            GROWTH FACTOR II ASSOCIATED (HUMAN);, mRNA sequence.
ACCESSION   AI016228
VERSION     AI016228.1  GI:3230564
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 373)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Trace considered overall poor quality
            Insert Length: 1829   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .373
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1623431"
                     /dev_stage="8-9 weeks"
                     /lab_host="DH10B"
                     /clone_lib="Soares_total_fetus_Nb2HF8_9w"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from pooled 8-9 week
                     (total) fetus material with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCTTAATTTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
ORIGIN

Alignment Scores:
Pred. No.:              2.12e+03        Length:         373
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AI016228 (1-373)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        265 CACAACCACAATCATAAC 248

RESULT 184
BG662150/c
LOCUS       BG662150                373 bp    mRNA    linear   EST 30-APR-2001
DEFINITION  Ljirnpest39-171-e12 Ljirnp Lambda HybriZap two-hybrid library Lotus
            corniculatus var. japonicus cDNA clone LP171-39-e12 5', mRNA
            sequence.
ACCESSION   BG662150
VERSION     BG662150.1  GI:13884072
KEYWORDS    EST.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1  (bases 1 to 373)
  AUTHORS   Poulsen,C. and Poedenphant,L.
  TITLE     Expressed sequence tags from Mesorhizobium loti infected roots of
            Lotus japonicus
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Poulsen,C.; Poedenphant,L.
            Laboratory of Gene Expression, Department of Molecular and
```

```
            Structural Biology
            University of Aarhus
            Gustav Wieds Vej 10C, DK-8000 Aarhus C, Denmark
            Tel: +45 89425007
            Fax: +45 86201222
            Email: CHP@mbio.aau.dk
            Seq primer: pADGAl4 5'-primer CCA CTA CAA TGG ATG ATG TAT ATA AC
            High quality sequence stop: 373.
FEATURES              Location/Qualifiers
     source           1. .373
                      /organism="Lotus corniculatus var. japonicus"
                      /mol_type="mRNA"
                      /cultivar="variety Gifu B-129"
                      /db_xref="taxon:34305"
                      /clone="LP171-39-e12"
                      /dev_stage="young plants"
                      /lab_host="Escherichia coli XL-1Blue MRF'"
                      /clone_lib="Ljirnp Lambda HybriZap two-hybrid library"
                      /note="Organ: whole roots and root nodule primordia;
                      Vector: pADGAL4; Site_1: EcoRI; Site_2: XhoI; Two times
                      polyA+ purified mRNA, oligodT primed, 5' adapter tag
                      GAATTCGGCACGAG "
ORIGIN

Alignment Scores:
Pred. No.:              2.12e+03        Length:         373
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-858-332G-28 (1-6) x BG662150 (1-373)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        286 CACAATCACAATCACAAT 269

RESULT 185
BI628630
LOCUS       BI628630                373 bp    mRNA    linear   EST 10-SEP-2001
DEFINITION  RH57130.5prime RH Drosophila melanogaster normalized Head pFlc-1
            Drosophila melanogaster cDNA clone RH57130 5 similar to CG14732:
            FBan0014732 GO:[] located on: 3R 87A6-87A6;: 08/22/2001, mRNA
            sequence.
ACCESSION   BI628630
VERSION     BI628630.1  GI:15530840
KEYWORDS    EST.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 373)
  AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B.,
            Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E.,
            George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G.,
            Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
            Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and
            Rubin,G.M.
  TITLE     BDGP/HHMI RH Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003693: arm:3R [7544898,7771380]
            estimated-cyto:86F7-87A8: 08/22/2001
            Plate: RH.571  row: C  column: 6
            High quality sequence stop: 256.
FEATURES              Location/Qualifiers
     source           1. .373
                      /organism="Drosophila melanogaster"
                      /mol_type="mRNA"
                      /db_xref="taxon:7227"
                      /clone="RH57130"
                      /sex="male and female"
                      /dev_stage="Adult"
                      /lab_host="DH5-alpha TonA"
                      /clone_lib="RH Drosophila melanogaster normalized Head
                      pFlc-1"
                      /note="Organ: head; Vector: pFlc1; Site_1: XhoI; Site_2:
                      BamHI; Library was kindly generated by Piero Carninci at
                      the RIKEN. The library  was normalized and excised using
                      Cre recombinase. Plasmid cDNA library."
ORIGIN

Alignment Scores:
Pred. No.:              2.12e+03        Length:         373
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-858-332G-28 (1-6) x BI628630 (1-373)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        159 CACAACCATAATCACAAT 176

RESULT 186
BY168107
LOCUS       BY168107                373 bp    mRNA    linear   EST 10-DEC-2002
DEFINITION  BY168107 RIKEN full-length enriched, bone marrow macrophage Mus
            musculus cDNA clone I830070B12 5', mRNA sequence.
ACCESSION   BY168107
VERSION     BY168107.1  GI:26304753
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 373)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
            Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
            Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
            Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
            Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
            Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
            Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
            Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
            Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
            Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
            Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
            Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
            Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
            Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
            Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
            Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
            Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
            Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
            Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
            Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
            Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
            Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
            Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
            Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
            Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
            Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
```

```
  JOURNAL   Nature 420, 563-573 (2002)
   PUBMED   12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
            Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
            Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,
            Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,
            Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and
            Hayashizaki,Y. Direct Submission
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
             cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
             Tissues were provided by David A. Hume ( Depts. of Biochemistry
            and Microbiology/Parasitology Institute for Molecular Bioscience
            University of Queensland Brisbane,Q 4072 Australia ) whose
            assistance we gratefully acknowledge.
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .373
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="I830070B12"
                     /tissue_type="bone marrow"
                     /cell_type="macrophage"
                     /clone_lib="RIKEN full-length enriched, bone marrow
                     macrophage"
ORIGIN

Alignment Scores:
Pred. No.:              2.12e+03        Length:         373
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     5               Gaps:           0

US-09-858-332G-28 (1-6) x BY168107 (1-373)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        275 CACAACCATAACCACAAC 292

RESULT 187
BZ545747
LOCUS       BZ545747                373 bp    DNA     linear   GSS 16-DEC-2002
DEFINITION  OGAFP26TC ZM2_0.7_1.5_KB Zea mays genomic clone ZMMBMa0050E04,
            genomic survey sequence.
ACCESSION   BZ545747
VERSION     BZ545747.1  GI:27099569
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 373)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
            Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
  TITLE     Consortium for Maize Genomics
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: methylation filtered.
FEATURES             Location/Qualifiers
     source          1. .373
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBMa0050E04"
                     /clone_lib="ZM2_0.7_1.5_KB"
                     /note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb
                     methylation filtered genomic DNA library"
ORIGIN

Alignment Scores:
Pred. No.:              2.12e+03        Length:         373
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     9               Gaps:           0

US-09-858-332G-28 (1-6) x BZ545747 (1-373)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        268 CATAATCATAATCATAAT 285

RESULT 188
BI242911
LOCUS       BI242911                374 bp    mRNA    linear   EST 12-JUL-2001
DEFINITION  RE40419.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
            Drosophila melanogaster cDNA clone RE40419 5 similar to CG14732:
            FBan0014732 located on: 3R 87A6-87A6;: 05/12/2001, mRNA sequence.
ACCESSION   BI242911
VERSION     BI242911.1  GI:14711740
KEYWORDS    EST.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 374)
  AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B.,
            Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E.,
            George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G.,
            Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
            Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and
            Rubin,G.M.
  TITLE     BDGP/HHMI RE Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
```

```
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003693: arm:3R [7544898,7771380]
            estimated-cyto:86F7-87A8: 05/12/2001
            Plate: RE.404  row: B  column: 7
            High quality sequence stop: 295.
FEATURES             Location/Qualifiers
     source          1. .374
                     /organism="Drosophila melanogaster"
                     /mol_type="mRNA"
                     /db_xref="taxon:7227"
                     /clone="RE40419"
                     /sex="male and female"
                     /dev_stage="0-24 hours mixed stage embryonic"
                     /lab_host="DH5-alpha TonA"
                     /clone_lib="RE Drosophila melanogaster normalized Embryo
                     pFlc-1"
                     /note="Organ: embryo; Vector: pFlc1; Site_1: XhoI; Site_2:
                     BamHI; Library was kindly generated by Piero Carninci at
                     the RIKEN. The library  was normalized and excised using
                     Cre recombinase. Plasmid cDNA library."
ORIGIN

Alignment Scores:
Pred. No.:                2.12e+03        Length:         374
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       2               Gaps:           0

US-09-858-332G-28 (1-6) x BI242911 (1-374)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        192 CACAACCATAATCACAAT 209

RESULT 189
BI306132/c
LOCUS       BI306132                374 bp    mRNA    linear   EST 20-JUL-2001
DEFINITION  NL_3_H06 Drought stress (leaf) Oryza sativa (indica cultivar-group)
            cDNA clone NL_3_H06 3', mRNA sequence.
ACCESSION   BI306132
VERSION     BI306132.1  GI:14981454
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 374)
  AUTHORS   Reddy,A.R., Ramakrishna,W., Chandrasekhar,A., Nagabhushan,I.,
            Ravindrababu,P. and Bennetzen,J.L.
  TITLE     Novel EST enrichment with normalized cDNA libraries from drought
            stressed rice (Oryza sativa L.cv Nagina 22)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Reddy AR
            Department of Plant Sciences, School of Life Sciences
            University of Hyderabad
            P.O. Central University, Hyderabad-500 046, A.P, India
            Tel: 0091-40-3010265
            Fax: 0091-40-3010145
            Email: arjulsl@uohyd.ernet.in
            Insert Length: 374   Std Error: 0.00
            Plate: 3  row: H  column: 06
            Seq primer: GTAAAACGACGGCCAGTG.
FEATURES             Location/Qualifiers
     source          1. .374
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nagina 22"
                     /db_xref="taxon:39946"
                     /clone="NL_3_H06"
                     /tissue_type="Entire leaf tissue"
                     /dev_stage="35 day-old seedlings"
                     /clone_lib="Drought stress (leaf)"
                     /note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized
                     leaf cDNA Library from drought stressed seedlings "
ORIGIN

Alignment Scores:
Pred. No.:                2.12e+03        Length:         374
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       3               Gaps:           0

US-09-858-332G-28 (1-6) x BI306132 (1-374)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        360 CACAACCACAACCACAAC 343

RESULT 190
BY065744
LOCUS       BY065744                374 bp    mRNA    linear   EST 06-DEC-2002
DEFINITION  BY065744 RIKEN full-length enriched, 16 days embryo heart Mus
            musculus cDNA clone I920044A20 5', mRNA sequence.
ACCESSION   BY065744
VERSION     BY065744.1  GI:26169647
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 374)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
            Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
            Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
            Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
            Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
            Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
            Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
            Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
            Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
            Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
            Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
            Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
            Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
            Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
            Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
            Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
            Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
            Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
            Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
            Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
            Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
            Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
            Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
            Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
            Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
            Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
  PUBMED    12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
```

Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

```
FEATURES             Location/Qualifiers
     source          1. .374
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="I920044A20"
                     /tissue_type="heart"
                     /dev_stage="16 days embryo"
                     /clone_lib="RIKEN full-length enriched, 16 days embryo
                     heart"
ORIGIN

Alignment Scores:
Pred. No.:              2.12e+03        Length:         374
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     5               Gaps:           0

US-09-858-332G-28 (1-6) x BY065744 (1-374)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        295 CACAACCATAACCACAAC 312

RESULT 191
BZ881398/c
LOCUS       BZ881398                374 bp    DNA     linear   GSS 18-MAR-2003
DEFINITION  CH240_218E10.TJ CHORI-240 Bos taurus genomic clone CH240_218E10,
            genomic survey sequence.
ACCESSION   BZ881398
VERSION     BZ881398.1  GI:29108800
KEYWORDS    GSS.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 374)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,
            Shvartsbeyn,A., Gebregeorgis,E., Chen,D., Riggs,F., de Jong,P.,
            Crawford,A.M. and McEwan,J.C.
  TITLE     Bovine BAC End Sequences from Library CHORI-240
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the bovine BAC library CHORI-240
            (http://www.chori.org/bacpac/bovine240.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/ordering_information.htm). This work
            was undertaken as part of the International Bovine BAC Mapping
            Consortium (IBBMC) by AgResearch Ltd., New Zealand and The
            Institute of Genomic Research (TIGR), USA.
            Plate: 218  row: E  column: 10
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .374
                     /organism="Bos taurus"
                     /mol_type="genomic DNA"
                     /strain="breed: Hereford"
                     /db_xref="taxon:9913"
                     /clone="CH240_218E10"
                     /sex="Male"
                     /cell_type="Blood"
                     /clone_lib="CHORI-240"
                     /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                     Hereford bull L1 Domino 99375; CHORI-240 Bovine BAC
                     library (Male) produced by Pieter de Jong"
ORIGIN

Alignment Scores:
Pred. No.:              2.12e+03        Length:         374
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     9               Gaps:           0

US-09-858-332G-28 (1-6) x BZ881398 (1-374)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         72 CATAATCATAATCATAAC 55

RESULT 192
AI109054
LOCUS       AI109054                375 bp    mRNA    linear   EST 19-APR-2001
DEFINITION  GH07561.5prime GH Drosophila melanogaster head pOT2 Drosophila
            melanogaster cDNA clone GH07561 5prime similar to X63731:
            D.melanogaster regulatory signal sequence (clone scs), mRNA
            sequence.
ACCESSION   AI109054
VERSION     AI109054.1  GI:3476950
KEYWORDS    EST.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 375)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 75  row: F  column: 1
            High quality sequence stop: 283.
FEATURES             Location/Qualifiers
```

```
     source          1. .375
                     /organism="Drosophila melanogaster"
                     /mol_type="mRNA"
                     /db_xref="taxon:7227"
                     /clone="GH07561"
                     /sex="male and female"
                     /dev_stage="adult"
                     /lab_host="DH5 - alpha"
                     /clone_lib="GH Drosophila melanogaster head pOT2"
                     /note="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2.  Plasmid cDNA library."
ORIGIN

Alignment Scores:
Pred. No.:              2.13e+03        Length:         375
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AI109054 (1-375)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        185 CACAACCATAATCACAAT 202

RESULT 193
BJ439631/c
LOCUS       BJ439631                375 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ439631 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv41h04 3', mRNA sequence.
ACCESSION   BJ439631
VERSION     BJ439631.1  GI:19414353
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 375)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES              Location/Qualifiers
     source           1. .375
                      /organism="Dictyostelium discoideum"
                      /mol_type="mRNA"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /clone="ddv41h04"
                      /sex="mat A"
                      /dev_stage="Growth phase"
                      /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:              2.13e+03        Length:         375
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-858-332G-28 (1-6) x BJ439631 (1-375)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        249 CACAATCACAATCACAAC 232

RESULT 194
BP803155/c
LOCUS       BP803155                375 bp    mRNA    linear   EST 15-FEB-2005
DEFINITION  BP803155 RAFL14 Arabidopsis thaliana cDNA clone RAFL23-27-N21 5',
            mRNA sequence.
ACCESSION   BP803155
VERSION     BP803155.1  GI:59839234
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 375)
  AUTHORS   Seki,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M., Akiyama,K.,
            Iida,K., Enju,A., Sakurai,T., Arakawa,T., Carninci,P., Fukuda,S.,
            Iida,J., Kawai,J., Sasaki,D., Shiraki,T., Hayashizaki,Y. and
            Shinozaki,K.
  TITLE     Large-Scale Analysis of RIKEN Arabidopsis Full-length cDNAs
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            An Arabidopsis full-length cDNA library was constructed essentially
            as reported previously (Seki et al., 1998, 2002).This clone is in a
            modified pBluescript vector.
            Please visit our web site (http://pfgweb.gsc.riken.jp and
            http://rarge.gsc.riken.jp) for further details.
FEATURES              Location/Qualifiers
     source           1. .375
                      /organism="Arabidopsis thaliana"
                      /mol_type="mRNA"
                      /db_xref="taxon:3702"
                      /clone="RAFL23-27-N21"
                      /tissue_type="root"
                      /lab_host="DH10B"
                      /clone_lib="RAFL14"
                      /note="Site_1: BamHI; Site_2: SalI"
ORIGIN

Alignment Scores:
Pred. No.:              2.13e+03        Length:         375
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-858-332G-28 (1-6) x BP803155 (1-375)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        112 CACAATCACAATCACAAT 95

RESULT 195
BQ120829/c
LOCUS       BQ120829                375 bp    mRNA    linear   EST 07-MAR-2003
DEFINITION  EST606405 mixed potato tissues Solanum tuberosum cDNA clone STMET23
            5' end, mRNA sequence.
ACCESSION   BQ120829
VERSION     BQ120829.2  GI:21920178
KEYWORDS    EST.
SOURCE      Solanum tuberosum (potato)
  ORGANISM  Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
```

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; lamiids; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 375)
AUTHORS Buell,C.R., Hart,A., Baker,B., Tanksley,S., Fry,W., Smart,C., Restrepo,S., Griffiths,H., van der Hoeven,R., Tsai,J. and Karamycheva,S.A.
TITLE Generation of a set of potato cDNA clones for microarray analyses
JOURNAL Unpublished (2002)
COMMENT On Apr 17, 2002 this sequence version replaced gi:20172791.
Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato-array@tigr.org
This clone can be obtained from the University of Arizona Genomics Institute. Orders can be made through URL:
http://genome.arizona.edu/orders/
Seq primer: T3.
FEATURES Location/Qualifiers
source 1. .375
/organism="Solanum tuberosum"
/mol_type="mRNA"
/cultivar="Kennebec or Binjte"
/db_xref="taxon:4113"
/clone="STMET23"
/tissue_type="mixed tissues"
/lab_host="SOLR"
/clone_lib="mixed potato tissues"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; supplier: Combination of untreated and Phytophthora infestans-treated libraries of stolons, leaves, leaflets, axillary buds of stem explants, petioles, germinating eyes, tubers, or roots."
ORIGIN

Alignment Scores:

Pred. No.:	2.13e+03	Length:	375
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	3	Gaps:	0

US-09-858-332G-28 (1-6) x BQ120829 (1-375)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        210 CATAACCATAACCATAAC 193
```

RESULT 196
BF695474
LOCUS BF695474 376 bp mRNA linear EST 22-DEC-2000
DEFINITION 601851868F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4075816 5', mRNA sequence.
ACCESSION BF695474
VERSION BF695474.1 GI:11980882
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 376)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM928 row: g column: 17
High quality sequence stop: 205.
FEATURES Location/Qualifiers
source 1. .376
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4075816"
/tissue_type="primitive neuroectoderm"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_56"
/note="Organ: brain; Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."
ORIGIN

Alignment Scores:

Pred. No.:	2.14e+03	Length:	376
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-28 (1-6) x BF695474 (1-376)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        261 CACAACCACAACCACAAC 278
```

RESULT 197
BP866389/c
LOCUS BP866389 376 bp mRNA linear EST 17-FEB-2005
DEFINITION BP866389 RAFL21 Arabidopsis thaliana cDNA clone RAFL21-77-I23 5', mRNA sequence.
ACCESSION BP866389
VERSION BP866389.1 GI:59927362
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 376)
AUTHORS Seki,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M., Akiyama,K., Iida,K., Enju,A., Sakurai,T., Arakawa,T., Carninci,P., Fukuda,S., Iida,J., Kawai,J., Sasaki,D., Shiraki,T., Hayashizaki,Y. and Shinozaki,K.
TITLE Large-Scale Analysis of RIKEN Arabidopsis Full-length cDNAs
JOURNAL Unpublished (2005)
COMMENT Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@rtc.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998, 2002).This clone is in a modified pBluescript vector.
Please visit our web site (http://pfgweb.gsc.riken.jp and http://rarge.gsc.riken.jp) for further details.
FEATURES Location/Qualifiers
source 1. .376

```
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="RAFL21-77-I23"
                     /lab_host="DH10B"
                     /clone_lib="RAFL21"
                     /note="Site_1: BamHI; Site_2: SalI; Subtraction Library.
                     The sequence was obtained from samples subjected to
                     various stress and plant hormones-treated"
ORIGIN

Alignment Scores:
Pred. No.:                2.14e+03        Length:         376
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       3               Gaps:           0

US-09-858-332G-28 (1-6) x BP866389 (1-376)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        143 CACAATCACAATCACAAT 126

RESULT 198
BP073663
LOCUS       BP073663                  377 bp    mRNA    linear   EST 27-AUG-2004
DEFINITION  BP073663 Lotus corniculatus var. japonicus nodulating root Lotus
            corniculatus var. japonicus cDNA clone GNf085g08 3', mRNA sequence.
ACCESSION   BP073663
VERSION     BP073663.1  GI:45630363
KEYWORDS    EST.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1  (bases 1 to 377)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Characteristics of the Lotus japonicus gene repertoire deduced from
            large-scale expressed sequence tag (EST) analysis
  JOURNAL   Plant Mol. Biol. 54 (3), 405-414 (2004)
   PUBMED   15284495
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .377
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="mRNA"
                     /isolate="Gifu B-129"
                     /db_xref="taxon:34305"
                     /clone="GNf085g08"
                     /tissue_type="nodulating root"
                     /clone_lib="Lotus corniculatus var. japonicus nodulating
                     root"
                     /note="root of hypernodulating mutant har1-3, rich in
                     nodule primordia"
ORIGIN

Alignment Scores:
Pred. No.:                2.14e+03        Length:         377
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       3               Gaps:           0

US-09-858-332G-28 (1-6) x BP073663 (1-377)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        290 CATAATCATAATCATAAT 307

RESULT 199
BQ982016
LOCUS       BQ982016                  377 bp    mRNA    linear   EST 21-AUG-2002
DEFINITION  QGE15F05.yg.ab1 QG_EFGHJ lettuce serriola Lactuca sativa cDNA clone
            QGE15F05, mRNA sequence.
ACCESSION   BQ982016
VERSION     BQ982016.1  GI:22399541
KEYWORDS    EST.
SOURCE      Lactuca sativa
  ORGANISM  Lactuca sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
            Cichorieae; Lactuca.
REFERENCE   1  (bases 1 to 377)
  AUTHORS   Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J.,
            Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y.,
            Lai,Z., Church,S., Jackson,L. and Bradford,K.
  TITLE     Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
            belongs to contig QG_CA_Contig3158, see http://cgpdb.ucdavis.edu/
            for details.
            Plate: QGE15  row: F  column: 05.
FEATURES             Location/Qualifiers
     source          1. .377
                     /organism="Lactuca sativa"
                     /mol_type="mRNA"
                     /cultivar="L.serriola"
                     /db_xref="taxon:4236"
                     /clone="QGE15F05"
                     /lab_host="E.coli"
                     /clone_lib="QG_EFGHJ lettuce serriola"
                     /note="Vector: pBRcDNASfiAB; The library was constructed
                     from 10 different sources of RNA from a single genotype.
                     Separate cDNAs were generated using primers that
                     incorporated unique 5' and 3' tags to distinguish each
                     source of RNA. cDNAs were then pooled, size-fractionated,
                     directionally cloned into a custom medium-copy vector and
                     transformations made with four size classes to minimize
                     size bias. Details of each source of RNA and library
                     construction can be obtained at http://cgpdb.ucdavis.edu/
                     TAG_TISSUE=roots
                     TAG_LIB=QG_EFGHJ lettuce serriola
                     TAG_SEQ=GTTGCACGGG"
ORIGIN

Alignment Scores:
Pred. No.:                2.14e+03        Length:         377
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       5               Gaps:           0

US-09-858-332G-28 (1-6) x BQ982016 (1-377)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         59 CACAACCACAACCACAAC 76
```

```
RESULT 200
CF369358/c
LOCUS       CF369358               377 bp    mRNA    linear   EST 27-AUG-2003
DEFINITION  rg58g06.y1 Meloidogyne hapla female SMART pGEM Meloidogyne hapla
            cDNA 5', mRNA sequence.
ACCESSION   CF369358
VERSION     CF369358.1  GI:34316604
KEYWORDS    EST.
SOURCE      Meloidogyne hapla
  ORGANISM  Meloidogyne hapla
            Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
            Tylenchoidea; Meloidogynidae; Meloidogyninae; Meloidogyne.
REFERENCE   1  (bases 1 to 377)
  AUTHORS   McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J.,
            Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B.,
            Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C.,
            Tsagareishvili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C.,
            Underwood,K., Steptoe,M., Allen,M., Person,B., Swaller,T.,
            Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     The Washington Univ. Nematode EST Project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: McCarter JP
            The Washington Univ. Nematode EST Project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Cloned unidirectionally. Poly(A)+ RNA was concentrated and purified
            using Dynabeads (Dynal) and mRNA eluted for first strand synthesis.
            First strand cDNA was created using MMLV RT (Powerscript, Clontech)
            and primed with oligo(dT) with XhoI site and 5'SMART 'anchor' added
            using chimeric DNA-RNA oligo. 12 PCR cycles were done using first
            strand and primers specific to SMART oligo and 3' end. Double
            stranded cDNA was digested using XhoI/NotI, fractioned on
            Chroma-spin 400 columns (Clontech) and ligated to digested
            pGEM-11zf(+) plasmid. Chemically competent DH10B cells were used as
            host cells. Library materials provided by Dr. David Bird of North
            Carolina State University. Library construction by Jeff Rousch. See
            www.nematode.net for additional project information.
            Putative full length read
            The vector to vector length is 378
            Seq primer: Sp6.
FEATURES             Location/Qualifiers
     source          1. .377
                     /organism="Meloidogyne hapla"
                     /mol_type="mRNA"
                     /db_xref="taxon:6305"
                     /tissue_type="whole organism"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="Meloidogyne hapla female SMART pGEM"
                     /note="Vector: plasmid (ampicillin resistant); Site_1:
                     XhoI; Site_2: NotI; Cloned unidirectionally. Poly(A)+ RNA
                     was concentrated and purified using Dynabeads (Dynal) and
                     mRNA eluted for first strand synthesis. First strand cDNA
                     was created using MMLV RT (Powerscript, Clontech) and
                     primed with oligo(dT) with XhoI site and 5'SMART 'anchor'
                     added using chimeric DNA-RNA oligo. 12 PCR cycles were
                     done using first strand and primers specific to SMART
                     oligo and 3' end. Double stranded cDNA was digested using
                     XhoI/NotI, fractioned on Chroma-spin 400 columns
                     (Clontech) and ligated to digested pGEM-11zf(+) plasmid.
                     Chemically competent DH10B cells were used as host cells.
                     Library materials provided by Dr. David Bird of North
                     Carolina State University. Library construction by Jeff
                     Rousch. See www.nematode.net for additional project
                     information."
ORIGIN

Alignment Scores:

Pred. No.:                2.14e+03      Length:          377
Score:                    42.00         Matches:         6
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       6             Gaps:            0

US-09-858-332G-28 (1-6) x CF369358 (1-377)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        204 CATAATCACAATCATAAT 187

RESULT 201
BY079363
LOCUS       BY079363               378 bp    mRNA    linear   EST 06-DEC-2002
DEFINITION  BY079363 RIKEN full-length enriched, pooled tissues, adult spleen,
            etc. Mus musculus cDNA clone K630018N14 5', mRNA sequence.
ACCESSION   BY079363
VERSION     BY079363.1  GI:26180542
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 378)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
            Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
            Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
            Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
            Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
            Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
            Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
            Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
            Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
            Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
            Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
            Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
            Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
            Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
            Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
            Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
            Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
            Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
            Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
            Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
            Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
            Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
            Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
            Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
            Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
            Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
   PUBMED   12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
            Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
            Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,
            Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,
            Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and
            Hayashizaki,Y. Direct Submission
             Computational Analysis of Full-Length Mouse cDNAs Compared with
```

Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

FEATURES Location/Qualifiers
source 1. .378
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="K630018N14"
/clone_lib="RIKEN full-length enriched, pooled tissues, adult spleen, etc."
/note="(dev_stage=adult,tissue_type=spleen,sex=male), (dev_stage=adult,tissue_type=kidney,sex=male), (dev_stage=adult,tissue_type=testis,sex=male), (dev_stage=adult,tissue_type=thymus,sex=male), (dev_stage=adult,tissue_type=heart,sex=male), (dev_stage=adult,tissue_type=colon,sex=male), (dev_stage=adult,tissue_type=stomach,sex=male), (dev_stage=adult,tissue_type=liver,sex=male), (dev_stage=13 days embryo,tissue_type=whole body,sex=mix),(dev_stage=14 days embryo,tissue_type=whole body,sex=mix),(dev_stage=16 days embryo,tissue_type=whole body,sex=mix),(dev_stage=17 days embryo,tissue_type=whole body,sex=mix),(dev_stage=15 days pregnant, adult,tissue_type=amnion,sex=female),(dev_stage=10 days neonate,tissue_type=brain,sex=mix),(dev_stage=10 days neonate,tissue_type=thymus,sex=mix),(dev_stage=10 days neonate,tissue_type=heart,sex=mix)"

ORIGIN

Alignment Scores:

Pred. No.:	2.15e+03	Length:	378
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	5	Gaps:	0

US-09-858-332G-28 (1-6) x BY079363 (1-378)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        295 CACAACCATAACCACAAC 312
```

RESULT 202
BQ583906
LOCUS BQ583906 379 bp mRNA linear EST 06-DEC-2002
DEFINITION E011887-024-004-L02-SP6 MPIZ-ADIS-024-inflorescence Beta vulgaris cDNA clone 024-004-L02 5-PRIME, mRNA sequence.
ACCESSION BQ583906
VERSION BQ583906.1 GI:26113483
KEYWORDS EST.
SOURCE Beta vulgaris
ORGANISM Beta vulgaris
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; Caryophyllales; Amaranthaceae; Beta.
REFERENCE 1 (bases 1 to 379)
AUTHORS Herwig,R., Schulz,B., Weisshaar,B., Hennig,S., Steinfath,M., Drungowski,M., Stahl,D., Wruck,W., Menze,A., O'Brien,J., Lehrach,H. and Radelof,U.
TITLE Construction of a 'unigene' cDNA clone set by oligonucleotide fingerprinting allows access to 25 000 potential sugar beet genes
JOURNAL Plant J. 32 (5), 845-857 (2002)
PUBMED 12472698
COMMENT Contact: Weisshaar B
ADIS DNA core facility at MPIZ
Max-Planck-Institute for Plant Breeding Research
Carl-von-Linne Weg 10, 50829 Koeln, Germany
Fax: 00492215062851
Email: weisshaa@mpiz-koeln.mpg.de
Insert Length: 379 Std Error: 0.00
Plate: 4 row: L column: 02
Seq primer: SP6; CATACGATTTAGGTGACACTATAG.

FEATURES Location/Qualifiers
source 1. .379
/organism="Beta vulgaris"
/mol_type="mRNA"
/cultivar="KWS2320 (double haploid, monogerm breeding line)"
/db_xref="GABI:182793"
/db_xref="taxon:161934"
/clone="024-004-L02"
/tissue_type="inflorescence"
/lab_host="EMDH10B"
/clone_lib="MPIZ-ADIS-024-inflorescence"
/note="Vector: pCMVSPORT6; Site_1: SalI; Site_2: NotI; cDNA library from sugar beet, library provided by KWS Kleinwanzlebener Saatzucht AG Einbeck, Germany, contact: b.schulz@kws.de; cloning sites SalI-NotI, primer sites and orientation: SP6-SalI-CCACGCGTCCG-5prime-cDNA-polyA-CC-NotI-T7; Note: Sequencing granted in the context of the GABI-Beet project, local PI: Dr. Katharina Schneider, coordinator: Prof. Christian Jung; Sequence submission managed by RZPD/GABI-Primary database: http://gabi.rzpd.de"

ORIGIN

Alignment Scores:

Pred. No.:	2.15e+03	Length:	379
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	5	Gaps:	0

US-09-858-332G-28 (1-6) x BQ583906 (1-379)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        150 CATAATCATAATCATAAT 167
```

RESULT 203
BP078874
LOCUS BP078874 380 bp mRNA linear EST 27-AUG-2004
DEFINITION BP078874 Lotus japonicus roots Lotus corniculatus var. japonicus cDNA clone MR037f01_f 3', mRNA sequence.
ACCESSION BP078874
VERSION BP078874.1 GI:45635535
KEYWORDS EST.
SOURCE Lotus corniculatus var. japonicus (Lotus japonicus)
ORGANISM Lotus corniculatus var. japonicus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae; Lotus.
REFERENCE 1 (bases 1 to 380)
AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
TITLE Characteristics of the Lotus japonicus gene repertoire deduced from large-scale expressed sequence tag (EST) analysis
JOURNAL Plant Mol. Biol. 54 (3), 405-414 (2004)
PUBMED 15284495

```
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .380
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="mRNA"
                     /isolate="Miyakojima MG-20"
                     /db_xref="taxon:34305"
                     /clone="MR037f01_f"
                     /tissue_type="roots"
                     /clone_lib="Lotus japonicus roots"
ORIGIN

Alignment Scores:
Pred. No.:              2.16e+03        Length:         380
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-858-332G-28 (1-6) x BP078874 (1-380)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        208 CATAATCATAATCATAAT 225

RESULT 204
CN572644
LOCUS       CN572644                380 bp    mRNA    linear   EST 04-MAY-2004
DEFINITION  rg90a05.x1 Meloidogyne hapla female SMART pGEM Meloidogyne hapla
            cDNA 3' similar to contains element MER30 repetitive element ;,
            mRNA sequence.
ACCESSION   CN572644
VERSION     CN572644.1  GI:46983332
KEYWORDS    EST.
SOURCE      Meloidogyne hapla
  ORGANISM  Meloidogyne hapla
            Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
            Tylenchoidea; Meloidogynidae; Meloidogyninae; Meloidogyne.
REFERENCE   1  (bases 1 to 380)
  AUTHORS   McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J.,
            Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B.,
            Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C.,
            Tsagareishvili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C.,
            Underwood,K., Steptoe,M., Allen,M., Person,B., Swaller,T.,
            Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     The Washington Univ. Nematode EST Project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: rg90a05.y1
            Contact: McCarter JP
            The Washington Univ. Nematode EST Project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Cloned unidirectionally. Poly(A)+ RNA was concentrated and purified
            using Dynabeads (Dynal) and mRNA eluted for first strand synthesis.
            First strand cDNA was created using MMLV RT (Powerscript, Clontech)
            and primed with oligo(dT) with XhoI site and 5'SMART 'anchor' added
            using chimeric DNA-RNA oligo. 12 PCR cycles were done using first
            strand and primers specific to SMART oligo and 3' end. Double
            stranded cDNA was digested using XhoI/NotI, fractioned on
            Chroma-spin 400 columns (Clontech) and ligated to digested
            pGEM-11zf(+) plasmid. Chemically competent DH10B cells were used as
            host cells. Library materials provided by Dr. David Bird of North
            Carolina State University. Library construction by Jeff Rousch. See
            www.nematode.net for additional project information.
            Seq primer: Primer name ambiguous
            High quality sequence stop: 380.
FEATURES             Location/Qualifiers
     source          1. .380
                     /organism="Meloidogyne hapla"
                     /mol_type="mRNA"
                     /db_xref="taxon:6305"
                     /tissue_type="whole organism"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="Meloidogyne hapla female SMART pGEM"
                     /note="Vector: plasmid (ampicillin resistant); Site_1:
                     XhoI; Site_2: NotI; Cloned unidirectionally. Poly(A)+ RNA
                     was concentrated and purified using Dynabeads (Dynal) and
                     mRNA eluted for first strand synthesis. First strand cDNA
                     was created using MMLV RT (Powerscript, Clontech) and
                     primed with oligo(dT) with XhoI site and 5'SMART 'anchor'
                     added using chimeric DNA-RNA oligo. 12 PCR cycles were
                     done using first strand and primers specific to SMART
                     oligo and 3' end. Double stranded cDNA was digested using
                     XhoI/NotI, fractioned on Chroma-spin 400 columns
                     (Clontech) and ligated to digested pGEM-11zf(+) plasmid.
                     Chemically competent DH10B cells were used as host cells.
                     Library materials provided by Dr. David Bird of North
                     Carolina State University. Library construction by Jeff
                     Rousch. See www.nematode.net for additional project
                     information."
ORIGIN

Alignment Scores:
Pred. No.:              2.16e+03        Length:         380
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     7               Gaps:           0

US-09-858-332G-28 (1-6) x CN572644 (1-380)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        162 CATAATCACAATCATAAT 179

RESULT 205
CO472973
LOCUS       CO472973                381 bp    mRNA    linear   EST 29-JUN-2005
DEFINITION  GQ0045.B3_E02 GQ004: Non-lignified secondary xylem from mature
            trees Picea glauca cDNA clone GenomeQuebec_Id:GQ0045E02 5', mRNA
            sequence.
ACCESSION   CO472973
VERSION     CO472973.1  GI:50141867
KEYWORDS    EST.
SOURCE      Picea glauca (white spruce)
  ORGANISM  Picea glauca
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
REFERENCE   1  (bases 1 to 381)
  AUTHORS   Morency,M.-J., Cooke,J., Pavy,N., Parsons,L., Paule,C., Seguin,A.,
            Retzel,E., Butterfield,Y., Barber,S., Yang,G., Stott,J.,
            Siddiqui,A., Holt,R., Marra,M. and MacKay,J.
  TITLE     Arborea EST sequencing in Picea glauca (white spruce)
  JOURNAL   Unpublished (2004)
COMMENT     Contact: John MacKay
            Centre de Recherche en Biologie Forestiere
            Universite Laval
            Pavillon Charles-Eugene Marchand, Quebec, Quebec, CANADA G1K 7P4
            Fax: 418 656 7493
            Email: jmackay@rsvs.ulaval.ca
            Center for Computational Genomics and Bioinformatics (CCGB),
            University of Minnesota, MN_id Identifier: MN5195507 Clone ID:
            GQ0045_E02 Clones available through: John MacKay, Ph. D. Professeur
            adjoint -Assistant professor EMAIL: jmackay@rsvs.ulaval.ca Centre
            de Recherche en Biologie Forestiere (Forest Biology Research
```

Center) Universite Laval Quebec, Quebec CANADA G1K 7P4
Plate: 5 row: 02 column: E
Seq primer: T3 Primer.

FEATURES Location/Qualifiers
source 1. .381
/organism="Picea glauca"
/mol_type="mRNA"
/strain="Two trees of provenance 5333 and one from 5206"
/db_xref="taxon:3330"
/clone="GenomeQuebec_Id:GQ0045E02"
/sex="Hermaphrodite"
/tissue_type="Non-lignified differentiating secondary xylem from normal vertical trees"
/dev_stage="Non-lignified secondary xylem from trees harvested 2.5 hours, 6 hours and 11 hours after day break"
/lab_host="E. coli DH10B cells"
/clone_lib="GQ004: Non-lignified secondary xylem from mature trees"
/note="Organ: Stem from ground to lower part of live crown, on 33 year old tree; Vector: pBluescript II SK (+) XR; Site_1: Eco-RI; Site_2: Xho-I; Tissue was harvested in mid-June, during formation of early wood. cDNA was prepared from 5 mg of poly A+ selected RNA and was directionally ligated into the pBluescript II SK (+) XR vector (Stratagene), transformed by electroporation into DH10B cells (In vitrogen) for propagation"

ORIGIN

Alignment Scores:

Pred. No.:	2.17e+03	Length:	381
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	7	Gaps:	0

US-09-858-332G-28 (1-6) x CO472973 (1-381)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         14 CACAATCACAACCACAAC 31
```

RESULT 206
W34128
LOCUS W34128 381 bp mRNA linear EST 12-SEP-1996
DEFINITION mb81c05.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone IMAGE:335816 5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION W34128
VERSION W34128.1 GI:1537490
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 381)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1316036.
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:217216
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 362.

FEATURES Location/Qualifiers
source 1. .381
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:335816"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares mouse p3NMF19.5"
/note="Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

ORIGIN

Alignment Scores:

Pred. No.:	2.17e+03	Length:	381
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	8	Gaps:	0

US-09-858-332G-28 (1-6) x W34128 (1-381)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        288 CACAACCATAACCACAAC 305
```

RESULT 207
BG588299/c
LOCUS BG588299 383 bp mRNA linear EST 12-APR-2001
DEFINITION EST490108 MHRP- Medicago truncatula cDNA clone pMHRP-47B9, mRNA sequence.
ACCESSION BG588299
VERSION BG588299.1 GI:13606439
KEYWORDS EST.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE 1 (bases 1 to 383)
AUTHORS Harrison,M.J., Liu,J., Town,C.D., Van Aken,S., Utterback,T., Cho,J. and Fraser,C.M.
TITLE ESTs from phosphate-starved roots of Medicago truncatula, 2001
JOURNAL Unpublished (2001)
COMMENT Contact: Harrison M.J.
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73401
Tel: 580-223-5810
Fax: 580-221-7380
Email: mjharrison@noble.org
The Samuel Roberts Noble Foundation: N266974e TIGR sequence name: MTHAW05TKB More information is available at: http://www.medicago.org
Seq primer: SKmod (CTA gAA CTA gtg gAT CC).

FEATURES Location/Qualifiers
source 1. .383
/organism="Medicago truncatula"

/mol_type="mRNA"
/cultivar="A17"
/db_xref="taxon:3880"
/clone="pMHRP-47B9"
/tissue_type="roots"
/dev_stage="phosphate-starved"
/lab_host="XLOLR"
/clone_lib="MHRP-"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; At the trifoliate stage, M. truncatula plants were transplanted to phosphate-free sand and grown for a further 30 days. During this period, they were fertilized twice weekly with 1/2 Hoaglands solutions containing 20uM potassium phosphate. cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into the Unizap XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in XLOLR cells."

ORIGIN

Alignment Scores:

Pred. No.:	2.18e+03	Length:	383
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-28 (1-6) x BG588299 (1-383)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        320 CATAATCATAATCATAAT 303
```

RESULT 208
BQ805889
LOCUS BQ805889 383 bp mRNA linear EST 31-JUL-2002
DEFINITION WHE3572_B08_C16ZS Wheat developing grains cDNA library Triticum aestivum cDNA clone WHE3572_B08_C16, mRNA sequence.
ACCESSION BQ805889
VERSION BQ805889.1 GI:22030098
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 383)
AUTHORS Altenbach,S., Anderson,O.D., Chao,S., Chin,A., Close,T.J., Cronin,K., Crossman,C., Fenton,R.D., Lazo,G.R., Pham,J., Rausch,C.J., Wilson,C. and Woo,J.
TITLE The structure and function of the expressed portion of the wheat genomes - Developing grains cDNA library
JOURNAL Unpublished (2002)
COMMENT Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oandersn@pw.usda.gov
Sequences have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
Seq primer: SK primer.
FEATURES Location/Qualifiers
source 1. .383
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Butte 86"
/db_xref="taxon:4565"
/clone="WHE3572_B08_C16"
/tissue_type="whole grains"
/dev_stage="3-44 days post anthesis seed"
/lab_host="E. coli SOLR"
/clone_lib="Wheat developing grains cDNA library"
/note="Vector: Lambda ZAP II, excised phagemid; Site_1: EcoRI; Plants were grown under six following different environmental regimes in greenhouse, Environment 1) 24oC/17oC day/night, well-watered, with post-anthesis fertilizer, Environment 2) 24oC/17oC day/night, well-watered, without post-anthesis fertilizer, Environment 3) 37oC/17oC day/night, well-watered, with post-anthesis fertilizer, Environment 4) 37oC/17oC day/night, well-watered, without post-anthesis fertilizer, Environment 5) 37oC/17oC day/night plus drought, with post-anthesis fertilizer, Environment 6) 37oC/17oC day/night plus drought, without post-anthesis fertilizer, developing wheat grains from the following were excised and frozen in liquid nitrogen, Environment 1 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44 DPA Environment 2 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44 DPA Environment 3 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 34 DPA Environment 4 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 34 DPA Environment 5 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 30 DPA Environment 6 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 30 DPA and total RNA was prepared by S. Altenbach and K. Cronin at USDA-ARS, Albany, CA. A cDNA library was made using poly (A) RNA, and the cDNA clones were in vivo excised to give pBluescript SK(-) phagemids in the TJ Close lab (Chin, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (others)."

ORIGIN

Alignment Scores:

Pred. No.:	2.18e+03	Length:	383
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	5	Gaps:	0

US-09-858-332G-28 (1-6) x BQ805889 (1-383)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        277 CACAACCACAACCACAAT 294
```

RESULT 209
BW690941/c
LOCUS BW690941 384 bp mRNA linear EST 27-APR-2005
DEFINITION BW690941 Lycopersicon esculentum cv. Micro-Tom fruit Lycopersicon esculentum cDNA clone FC20BD02 5', mRNA sequence.
ACCESSION BW690941
VERSION BW690941.1 GI:62933352
KEYWORDS EST.
SOURCE Lycopersicon esculentum (Solanum lycopersicum)
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 384)
AUTHORS Tsugane,T., Watanabe,M., Yano,K., Suzuki,H., Sakurai,N. and Shibata,D.
TITLE Expressed sequence tags of full-length cDNA clones prepared from the laboratory-grown miniature tomato (Lycopersicon esculentum) cultivar Micro-Tom
JOURNAL Unpublished (2005)
COMMENT Contact: Daisuke Shibata
Kazusa DNA Research Institute;
2-6-7 Kazusa-Kamatari, Kisarazu, Chiba, 292-0818, Japan
Tel: 81-438-52-3947
Fax: 81-438-52-3948

```
            Email: shibata@kazusa.or.jp.
FEATURES             Location/Qualifiers
     source          1. .384
                     /organism="Lycopersicon esculentum"
                     /mol_type="mRNA"
                     /cultivar="Micro-Tom"
                     /db_xref="taxon:4081"
                     /clone="FC20BD02"
                     /tissue_type="fruit"
                     /clone_lib="Lycopersicon esculentum cv. Micro-Tom fruit"
ORIGIN

Alignment Scores:
Pred. No.:                2.19e+03       Length:         384
Score:                    42.00          Matches:        6
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       5              Gaps:           0

US-09-858-332G-28 (1-6) x BW690941 (1-384)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        359 CATAACCATAACCATAAC 342

RESULT 210
BH883598
LOCUS       BH883598                384 bp    DNA     linear   GSS 05-AUG-2002
DEFINITION  hw45f09.g1 WGS-ZmaysF (JM107 adapted methyl filtered) Zea mays
            genomic clone hw45f09 5', genomic survey sequence.
ACCESSION   BH883598
VERSION     BH883598.1  GI:22119495
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 384)
  AUTHORS   Rabinowicz,P.D., O'Shaughnessy,A.L., Balija,V., Dedhia,N.,
            Katzenburger,F., King,L., Miller,B., Muller,S., Nascimento,L.,
            Zutavern,T., McCombie,W.R. and Martienssen,R.A.
  TITLE     Genomic shotgun sequences from Zea mays (methyl-filtered)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: W. Richard McCombie
            Lita Annenberg Hazen Genome Sequencing Center
            Cold Spring Harbor Laboratory
            PO Box 100, Cold Spring Harbor, NY 11724, USA
            Tel: 516 367 8884
            Fax: 516 367 8874
            Email: mccombie@cshl.org
            Plate: hw45  row: f  column: 09
            Seq primer: -21M13UnivFwd
            Class: shotgun
            High quality sequence stop: 384.
FEATURES             Location/Qualifiers
     source          1. .384
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="hw45f09"
                     /lab_host="JM107 or DH5a"
                     /clone_lib="WGS-ZmaysF (JM107 adapted methyl filtered)"
                     /note="Organ: immature ears; Site_1: Xba I; Site_2: Xba I;
                     The vector was digested with XbaI and one nucleotide was
                     added by  fill in in the recessive 3' end. The genomic DNA
                     was nebulized, end  repaired, adaptor ligated and size
                     fractionated using sephadex. The  resulting fragments were
                     between 0.8 and 3 kb and were cloned into  the vector
                     (.x/y reads in M13mp19, .b/g reads in pUC19). The same
                     ligation was transformed in either JM107 or DH5a. "
ORIGIN

Alignment Scores:
Pred. No.:                2.19e+03       Length:         384
Score:                    42.00          Matches:        6
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       9              Gaps:           0

US-09-858-332G-28 (1-6) x BH883598 (1-384)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        104 CATAACCATAACCATAAC 121

RESULT 211
CB826676
LOCUS       CB826676                385 bp    mRNA    linear   EST 16-MAY-2003
DEFINITION  LjNEST65f4r Lotus japonicus nodule library 5 and 7 week-old Lotus
            corniculatus var. japonicus cDNA 5', mRNA sequence.
ACCESSION   CB826676
VERSION     CB826676.1  GI:29964904
KEYWORDS    EST.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1  (bases 1 to 385)
  AUTHORS   Colebatch,G., Freund,S., Trevaskis,B and Udvardi,M.
  TITLE     Lotus japonicus root nodule ESTs: tools for functional genomics
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Udvardi MK
            Molecular Plant Nutrition
            Max Planck Institute of Molecular Plant Physiology
            Am Muehlenberg 1, 14476 Golm, Germany
            Fax: 49 331 567 8250
            Email: udvardi@mpimp-golm.mpg.de
            Seq primer: T7
            High quality sequence stop: 385.
FEATURES             Location/Qualifiers
     source          1. .385
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="mRNA"
                     /cultivar="Gifu (B-129)"
                     /db_xref="taxon:34305"
                     /dev_stage="5 and 7 week-old plants"
                     /clone_lib="Lotus japonicus nodule library 5 and 7
                     week-old"
                     /note="Organ: Nodule; Vector: pSPORT1; Site_1: SalI;
                     Site_2: NotI; The library was prepared using mRNA
                     extracted from nodules of 5 and 7 week-old Lotus plants.
                     Nodules were induced by, and contained Mesorhizobium
                     strain R7A."
ORIGIN

Alignment Scores:
Pred. No.:                2.19e+03       Length:         385
Score:                    42.00          Matches:        6
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       6              Gaps:           0

US-09-858-332G-28 (1-6) x CB826676 (1-385)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        161 CATAATCATAACCATAAT 178

RESULT 212
```

```
BF068492
LOCUS       BF068492                386 bp    mRNA    linear   EST 13-JUL-2004
DEFINITION  st82f08.y1 Gm-c1054 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1054-1024 5' similar to SW:IF4Z_TOBAC Q40468 EUKARYOTIC
            INITIATION FACTOR 4A-15 ;, mRNA sequence.
ACCESSION   BF068492
VERSION     BF068492.1  GI:10845345
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 386)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,
            Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
            Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
            Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
            Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            When it has been determined, an EST from the other end of this
            clone is listed in the 'Other ESTs on clone' field. This clone is
            available through: Biogenetic Services, 801 32nd Ave. Brookings, SD
            57006 USA (phone: 800 423 4163; email:
            info@biogeneticservices.com).
FEATURES             Location/Qualifiers
     source          1. .386
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /cultivar="Harosoy"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1054-1024"
                     /tissue_type="Leaf, 3 week old, greenhouse grown"
                     /lab_host="DH10B"
                     /clone_lib="Gm-c1054"
                     /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                     XhoI; The Harosoy NIL was constructed and seed was
                     provided by Dr. J. Specht, University of Nebraska
                     (Shoemaker and Specht, 1995). The cDNA library was
                     constructed from mRNA isolated from leaf tissue at various
                     developmental stages of 3 week old greenhouse grown
                     plants. Complementary DNA was  synthesized from mRNA using
                     a primer consisting of a poly(dT) sequence  with a XhoI
                     restriction site and a 3' anchor. EcoRI adapters were
                     ligated  to the blunt-ended cDNA fragments followed by
                     XhoI digestion. The cDNA  fragments were directionally
                     cloned into the EcoRI-XhoI restriction site of  the
                     pBluescript vector. The ligated cDNA fragments were
                     transformed into  DH10B host cells (GibcoBRL). This
                     library was constructed in cooperation  with Dr. Paul
                     Keim's laboratory at Northern Arizona University."
ORIGIN

Alignment Scores:
Pred. No.:              2.2e+03        Length:          386
Score:                  42.00          Matches:         6
Percent Similarity:     100.00%        Conservative:    0
Best Local Similarity:  100.00%        Mismatches:      0
Query Match:            100.00%        Indels:          0
DB:                     2              Gaps:            0

US-09-858-332G-28 (1-6) x BF068492 (1-386)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         39 CACAACCACAACCACAAC 56

RESULT 213
Z26004
LOCUS       Z26004                  386 bp    mRNA    linear   EST 09-SEP-1993
DEFINITION  ATTS1281 Grenoble-B Arabidopsis thaliana cDNA clone GBGe119a, mRNA
            sequence.
ACCESSION   Z26004
VERSION     Z26004.1  GI:398292
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 386)
  AUTHORS   CNRS.
  TITLE     The Arabidopsis thaliana transcribed genome: the GDR cDNA program
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Mache R., Quigley F., Thomas F., Yu DY.
            CNRS URA 1178
            Laboratoire de Biologie Moleculaire vegetale
            B.P. 53X,38041 Grenoble Cedex,France
            Email: FQUIGLEY@cicg990.grenet.fr.
FEATURES             Location/Qualifiers
     source          1. .386
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /ecotype="C24"
                     /db_xref="taxon:3702"
                     /clone="GBGe119a"
                     /clone_lib="Grenoble-B"
                     /note="Vector: Lambda ZAP; tissue=Flower buds of
                     A.thaliana ecotype columbia C24; clone_library=Grenoble-B;
                     Cloning vector: Lambda ZAP."
ORIGIN

Alignment Scores:
Pred. No.:              2.2e+03        Length:          386
Score:                  42.00          Matches:         6
Percent Similarity:     100.00%        Conservative:    0
Best Local Similarity:  100.00%        Mismatches:      0
Query Match:            100.00%        Indels:          0
DB:                     8              Gaps:            0

US-09-858-332G-28 (1-6) x Z26004 (1-386)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         13 CACAACCACAACCACAAT 30

RESULT 214
AI442701
LOCUS       AI442701                387 bp    mRNA    linear   EST 24-JUL-2004
DEFINITION  sa85c07.y1 Gm-c1004 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1004-6085 5', mRNA sequence.
ACCESSION   AI442701
VERSION     AI442701.1  GI:4298019
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 387)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,
            Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
            Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
            Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
            Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
```

```
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: AW350275 corresponding to Gm-r1021-2392 (3')
            Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            When it has been determined, an EST from the other end of this
            clone is listed in the 'Other ESTs on clone' field. This clone is
            available through: Biogenetic Services, 801 32nd Ave. Brookings, SD
            57006 USA (phone: 800 423 4163; email: info@biogeneticservices.com)
            Insert Length: 599   Std Error: 0.00
            Seq primer: -40RP from Gibco
            High quality sequence stop: 364
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .387
                      /organism="Glycine max"
                      /mol_type="mRNA"
                      /cultivar="Williams"
                      /db_xref="taxon:3847"
                      /clone="GENOME SYSTEMS CLONE ID: Gm-c1004-6085"
                      /tissue_type="root"
                      /lab_host="XL10-Gold"
                      /clone_lib="Gm-c1004"
                      /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
                      XhoI; Root cDNA. The mRNA was isolated from entire roots
                      of 8 day old 'Williams' seedlings which were propagated on
                      paper towels with distilled water. Stratagene's  cDNA
                      Synthesis Kit (catalog #200401) was used to synthesize the
                      cDNA. First- strand synthesis was performed with 5-methyl
                      dCTP, hence the ligated cDNA is hemimethylated.
                      Stratagene's first-strand synthesis primer was used
                      [GAGAGAGAGAGAGAGAGAACTAGTCTCGAG(T)-18]. After
                      second-strand synthesis, the cDNA ends were 'polished'
                      with clone Pfu DNA polymerase, ligated to EcoRI adapters,
                      and phosphorylated. The XhoI site within the first-strand
                      synthesis primer was restricted by digestion with XhoI;
                      all XhoI sites in the  cDNA would be protected by their
                      hemimethylated status. The cDNA constructs were
                      size-fractionated with a 500bp cutoff, using GibcoBRL Life
                      Technologies' cDNA Size Fractionation column. The column
                      eluent was then ligated into  Stratagene's pBluescript II
                      XR Predigested vector (pBluescript II SK(+) that had been
                      digested with EcoRI and XhoI, and phosphorylated). Both
                      the white  and blue colonies appear to contain recombinant
                      plasmids with cDNA inserts. Blue colonies 9n=15) have been
                      sequenced, and possess putative cDNA inserts. This library
                      was constructed by Dr. Paul Keim & Virginia H. Coryell,
                      Department of Biology, Box5640, Northern Arizona
                      University, Flagstaff, AZ 86011, Phone: 520-523-1078 (Dr.
                      Paul Keim), 520-523-1372 (Virginia H. Coryell), Fax:
                      520-523-7500, email: paul.keim@nau.edu,
                      virginia.coryell@nau.edu"
ORIGIN

Alignment Scores:
Pred. No.:                2.2e+03        Length:          387
Score:                    42.00          Matches:         6
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       1              Gaps:            0

US-09-858-332G-28 (1-6) x AI442701 (1-387)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        290 CACAATCATAATCACAAT 307
```

```
RESULT 215
T03482
LOCUS       T03482                   387 bp    mRNA    linear   EST 14-FEB-1997
DEFINITION  IB36 Infant brain, Bento Soares Homo sapiens cDNA clone IB36 3'end,
            mRNA sequence.
ACCESSION   T03482
VERSION     T03482.1  GI:314722
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 387)
  AUTHORS   Khan,A.S., Wilcox,A.S., Polymeropoulos,M.H., Hopkins,J.A.,
            Stevens,T.J., Robinson,M., Orpana,A.K. and Sikela,J.M.
  TITLE     Single pass sequencing and physical and genetic mapping of human
            cDNAs
  JOURNAL   Nat. Genet. 2, 180-185 (1992)
   PUBMED   1345165
COMMENT     On Sep 21, 1992 this sequence version replaced gi:276139.
            Contact: Sikela JM
            Department of Pharmacology
            University of Colorado Health Sciences Center
            Box C236, 4200 E. 9th Ave, Denver CO 80262-0236
            Tel: 3032708637
            Fax: 3032707097
            Email: nikki@tally.uchsc.edu
            Seq primer: -21M13 Universal.
FEATURES              Location/Qualifiers
     source           1. .387
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="ATCC (inhost):86115"
                      /db_xref="GDB:D0S8690E"
                      /db_xref="taxon:9606"
                      /clone="IB36"
                      /lab_host="E. coli DH5-alpha"
                      /clone_lib="Infant brain, Bento Soares"
                      /note="Vector: BA, M13-derived; Site_1: HindIII; Site_2:
                      NotI; The infant brain library, constructed by Bento
                      Soares, Columbia University, was oligo-(dT) primed and
                      directionally cloned into an M13-derived plasmid using
                      total brain mRNA from a 72-day old human female afflicted
                      with spinal muscular atrophy. "
ORIGIN

Alignment Scores:
Pred. No.:                2.2e+03        Length:          387
Score:                    42.00          Matches:         6
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       8              Gaps:            0

US-09-858-332G-28 (1-6) x T03482 (1-387)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        191 CATAATCATAATCATAAT 208

RESULT 216
CW292826
LOCUS       CW292826                 388 bp    DNA     linear   GSS 31-OCT-2004
DEFINITION  104_773_11414297_116_35587_079 Sorghum methylation filtered library
            (LibID: 104) Sorghum bicolor genomic clone 11414297, genomic survey
            sequence.
ACCESSION   CW292826
VERSION     CW292826.1  GI:55009014
KEYWORDS    GSS.
SOURCE      Sorghum bicolor (sorghum)
  ORGANISM  Sorghum bicolor
```

```
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 388)
  AUTHORS   Bedell,J.A., Budiman,M.A., Nunberg,A., Citek,R.W., Robbins,D.,
            Jones,J., Flick,E., Rohlfing,T., Fries,J., Bradford,K.,
            McMenamy,J., Smith,M., Holeman,H., Roe,B.A, Wiley,G., Korf,I.F.,
            Rabinowicz,P.D., Lakey,N., McCombie,W.R., Jeddeloh,J.A. and
            Martienssen,R.A.
  TITLE     Sorghum genome sequencing by methylation filtration
  JOURNAL   PLoS Biol. 3 (1), e13 (2005)
   PUBMED   15660154
COMMENT     Contact: Bedell JA
            Orion Genomics, LLC
            4041 Forest Park Ave, St. Louis, MO 63108, USA
            Tel: 314 615 6979
            Fax: 314 615 5975
            Email: jbedell@oriongenomics.com
            Plate: 773  row: b  column: 17
            Seq primer: T3 Reverse
            Class: methylation filtered
            High quality sequence stop: 388.
FEATURES             Location/Qualifiers
     source          1. .388
                     /organism="Sorghum bicolor"
                     /mol_type="genomic DNA"
                     /cultivar="ATx623"
                     /db_xref="taxon:4558"
                     /clone="11414297"
                     /clone_lib="Sorghum methylation filtered library (LibID:
                     104)"
                     /note="Organ: leaf; Vector: pBCSK(-); Site_1: HincII; DNA
                     prepared from purified nuclei was randomly sheared,
                     end-repaired, size fractionated to enrich for the 0.5 to 5
                     kb fraction, ligated into HincII-digested pBCSK(-) vector
                     and electroporated into E. coli cells. This is a
                     methylation filtered library."
ORIGIN

Alignment Scores:
Pred. No.:              2.21e+03        Length:         388
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     10              Gaps:           0

US-09-858-332G-28 (1-6) x CW292826 (1-388)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         28 CATAATCATAATCATAAT 45

RESULT 217
AW912149
LOCUS       AW912149                389 bp    mRNA    linear   EST 25-MAY-2000
DEFINITION  uf42b08.y1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
            IMAGE:1514007 5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0
            KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION   AW912149
VERSION     AW912149.1  GI:8077746
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 389)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Other_ESTs: uf42b08.x1
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:940859
            Seq primer: -40RP from Gibco
            High quality sequence stop: 387.
FEATURES             Location/Qualifiers
     source          1. .389
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1514007"
                     /sex="female (lactating)"
                     /tissue_type="mammary gland"
                     /lab_host="DH10B"
                     /clone_lib="Soares_mammary_gland_NMLMG"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from mammary
                     gland tissue from a lactating female, and was then primed
                     with a Not I - oligo(dT) primer. Double-stranded  cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I  and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector.  Library is normalized.  Library
                     was constructed by Bento Soares and M. Fatima Bonaldo. "
ORIGIN

Alignment Scores:
Pred. No.:              2.22e+03        Length:         389
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AW912149 (1-389)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        117 CACAACCATAACCACAAC 134

RESULT 218
AW332717/c
LOCUS       AW332717                390 bp    mRNA    linear   EST 31-JAN-2000
DEFINITION  S12A5 AGS-1 Pneumocystis carinii cDNA 3', mRNA sequence.
ACCESSION   AW332717
VERSION     AW332717.1  GI:6829074
KEYWORDS    EST.
SOURCE      Pneumocystis carinii
  ORGANISM  Pneumocystis carinii
            Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;
            Pneumocystidaceae; Pneumocystis.
REFERENCE   1  (bases 1 to 390)
  AUTHORS   Smulian,A.G., Arnold,J., Weise,M., Wunderlich,J., Staben,C.,
            Edman,J.C., Kovacs,J. and Cushion,M.
  TITLE     Expressed sequence tags from Pneumocystis carinii
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Staben C
            School of Biological Sciences
            University of Kentucky
            101 Morgan Building, University of Kentucky, Lexington, KY
            40506-0225, USA
            Tel: 606 257 2161
            Fax: 606 257 1717
            Email: staben@pop.uky.edu.
FEATURES             Location/Qualifiers
     source          1. .390
                     /organism="Pneumocystis carinii"
                     /mol_type="mRNA"
                     /db_xref="taxon:4754"
                     /lab_host="E. coli"
                     /clone_lib="AGS-1"
                     /note="Vector: Lambda ZAP II; Site_1: EcoRI; Site_2: XhoI;
```

P. carinii organisms (3x10e9) from a single rat (99-1-6, sacrificed on 3/17/99) at Cincinnati VA facilities. Trizol extracted RNA. Oligo dT priming, standard conditions described by vendor, Stratagene. Further details see www.uky.edu/Project/Pneumocystis/"

ORIGIN

Alignment Scores:

Pred. No.:	2.22e+03	Length:	390
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0

US-09-858-332G-28 (1-6) x AW332717 (1-390)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        357 CACAATCACAATCACAAT 340
```

RESULT 219
BQ980435
LOCUS BQ980435 390 bp mRNA linear EST 21-AUG-2002
DEFINITION QGE11A15.yg.ab1 QG_EFGHJ lettuce serriola Lactuca sativa cDNA clone QGE11A15, mRNA sequence.
ACCESSION BQ980435
VERSION BQ980435.1 GI:22397958
KEYWORDS EST.
SOURCE Lactuca sativa
ORGANISM Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; campanulids; Asterales; Asteraceae; Cichorioideae; Cichorieae; Lactuca.
REFERENCE 1 (bases 1 to 390)
AUTHORS Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L., Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J., Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z., Church,S., Jackson,L. and Bradford,K.
TITLE Lettuce and Sunflower ESTs from the Compositae Genome Project http://compgenomics.ucdavis.edu/
JOURNAL Unpublished (2002)
COMMENT Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
belongs to contig QG_CA_Contig3158, see http://cgpdb.ucdavis.edu/ for details.
Plate: QGE11 row: A column: 15.

FEATURES Location/Qualifiers
source 1. .390
/organism="Lactuca sativa"
/mol_type="mRNA"
/cultivar="L.serriola"
/db_xref="taxon:4236"
/clone="QGE11A15"
/lab_host="E.coli"
/clone_lib="QG_EFGHJ lettuce serriola"
/note="Vector: pBRcDNASfiAB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at http://cgpdb.ucdavis.edu/
TAG_TISSUE=roots
TAG_LIB=QG_EFGHJ lettuce serriola
TAG_SEQ=GTTGCACGGG"

ORIGIN

Alignment Scores:

Pred. No.:	2.22e+03	Length:	390
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	5	Gaps:	0

US-09-858-332G-28 (1-6) x BQ980435 (1-390)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         60 CACAACCACAACCACAAC 77
```

RESULT 220
BY086496
LOCUS BY086496 390 bp mRNA linear EST 07-DEC-2002
DEFINITION BY086496 RIKEN full-length enriched, pooled tissues, adult spleen, etc. Mus musculus cDNA clone K630061E22 5', mRNA sequence.
ACCESSION BY086496
VERSION BY086496.1 GI:26196303
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 390)
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
PUBMED 12466851
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,

Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

FEATURES Location/Qualifiers
source 1. .390
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="K630061E22"
/clone_lib="RIKEN full-length enriched, pooled tissues, adult spleen, etc."
/note="(dev_stage=adult,tissue_type=spleen,sex=male), (dev_stage=adult,tissue_type=kidney,sex=male), (dev_stage=adult,tissue_type=testis,sex=male), (dev_stage=adult,tissue_type=thymus,sex=male), (dev_stage=adult,tissue_type=heart,sex=male), (dev_stage=adult,tissue_type=colon,sex=male), (dev_stage=adult,tissue_type=stomach,sex=male), (dev_stage=adult,tissue_type=liver,sex=male), (dev_stage=13 days embryo,tissue_type=whole body,sex=mix),(dev_stage=14 days embryo,tissue_type=whole body,sex=mix),(dev_stage=16 days embryo,tissue_type=whole body,sex=mix),(dev_stage=17 days embryo,tissue_type=whole body,sex=mix),(dev_stage=15 days pregnant, adult,tissue_type=amnion,sex=female),(dev_stage=10 days neonate,tissue_type=brain,sex=mix),(dev_stage=10 days neonate,tissue_type=thymus,sex=mix),(dev_stage=10 days neonate,tissue_type=heart,sex=mix)"

ORIGIN

Alignment Scores:

Pred. No.:	2.22e+03	Length:	390
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	5	Gaps:	0

US-09-858-332G-28 (1-6) x BY086496 (1-390)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        276 CACAACCATAACCACAAC 293
```

RESULT 221
BY077809
LOCUS BY077809 391 bp mRNA linear EST 06-DEC-2002
DEFINITION BY077809 RIKEN full-length enriched, pooled tissues, adult spleen, etc. Mus musculus cDNA clone K630011J24 5', mRNA sequence.
ACCESSION BY077809
VERSION BY077809.1 GI:26179207
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 391)
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
PUBMED 12466851
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

FEATURES Location/Qualifiers
source 1. .391
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="K630011J24"
/clone_lib="RIKEN full-length enriched, pooled tissues,

adult spleen, etc."
/note="(dev_stage=adult,tissue_type=spleen,sex=male),
(dev_stage=adult,tissue_type=kidney,sex=male),
(dev_stage=adult,tissue_type=testis,sex=male),
(dev_stage=adult,tissue_type=thymus,sex=male),
(dev_stage=adult,tissue_type=heart,sex=male),
(dev_stage=adult,tissue_type=colon,sex=male),
(dev_stage=adult,tissue_type=stomach,sex=male),
(dev_stage=adult,tissue_type=liver,sex=male),
(dev_stage=13 days embryo,tissue_type=whole
body,sex=mix),(dev_stage=14 days embryo,tissue_type=whole
body,sex=mix),(dev_stage=16 days embryo,tissue_type=whole
body,sex=mix),(dev_stage=17 days embryo,tissue_type=whole
body,sex=mix),(dev_stage=15 days pregnant,
adult,tissue_type=amnion,sex=female),(dev_stage=10 days
neonate,tissue_type=brain,sex=mix),(dev_stage=10 days
neonate,tissue_type=thymus,sex=mix),(dev_stage=10 days
neonate,tissue_type=heart,sex=mix)"
ORIGIN

Alignment Scores:

Pred. No.:	2.23e+03	Length:	391
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	5	Gaps:	0

US-09-858-332G-28 (1-6) x BY077809 (1-391)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        284 CACAACCATAACCACAAC 301
```

RESULT 222
W96855
LOCUS W96855 391 bp mRNA linear EST 16-JUL-1996
DEFINITION mf95e07.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA clone IMAGE:422052 5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION W96855
VERSION W96855.1 GI:1427340
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 391)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:256604
Seq primer: ETPrimer
High quality sequence stop: 342.
FEATURES Location/Qualifiers
source 1. .391
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:422052"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/clone_lib="Soares mouse embryo NbME13.5 14.5"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTTT T 3'], on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo. "
ORIGIN

Alignment Scores:

Pred. No.:	2.23e+03	Length:	391
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	8	Gaps:	0

US-09-858-332G-28 (1-6) x W96855 (1-391)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        295 CACAACCATAACCACAAC 312
```

RESULT 223
CC153019/c
LOCUS CC153019 391 bp DNA linear GSS 25-APR-2003
DEFINITION CSU-K34.126N10.T7 CSU-K34 Aedes aegypti genomic clone CSU-K34-126N10, genomic survey sequence.
ACCESSION CC153019
VERSION CC153019.1 GI:30106315
KEYWORDS GSS.
SOURCE Aedes aegypti (yellow fever mosquito)
ORGANISM Aedes aegypti
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae; Culicinae; Culicini; Aedes; Stegomyia.
REFERENCE 1 (bases 1 to 391)
AUTHORS Loftus,B., Shetty,J., Knudson,D. and Severson,D.
TITLE BAC end sequencing of Aedes aegypti
JOURNAL Unpublished (2003)
COMMENT Other_GSSs: CSU-K34.126N10.SP6
Contact: Brendan Loftus
Department of Eukaryotic Genomics
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-3543
Fax: 301-838-0208
Email: enta@tigr.org
Library was provided by Susan Brown and Dennis Knudson at Colorado State University.
Seq primer: T7
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .391
/organism="Aedes aegypti"
/mol_type="genomic DNA"
/db_xref="taxon:7159"
/clone="CSU-K34-126N10"
/clone_lib="CSU-K34"
/note="Vector: pBACe3.6; Site_1: EcoRI; Source DNA: Aedes aegypti; strain unknown [derived from freshly hatched

```
                larvae at the Virus Research Centre, Poona, India.
                Reference: SINGH, K. R. P., 1967 Cell cultures derived
                from larvae of Aedes albopictus (Skuse) and Aedes aegypti
                (L.). Current Science 36: 506-508]; ATC-10 cell line ATCC
                CCL-125"
ORIGIN

Alignment Scores:
Pred. No.:              2.23e+03        Length:         391
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     9               Gaps:           0

US-09-858-332G-28 (1-6) x CC153019 (1-391)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        381 CACAATCACAATCACAAT 364

RESULT 224
BE637465
LOCUS       BE637465                392 bp    mRNA    linear   EST 25-AUG-2000
DEFINITION  WHE0859_E11_I21ZS Wheat 20-45 DAP spike cDNA library Triticum
            aestivum cDNA clone WHE0859_E11_I21, mRNA sequence.
ACCESSION   BE637465
VERSION     BE637465.1  GI:9920576
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 392)
  AUTHORS   Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D.,
            Han,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J.,
            Seaton,C.L. and Tong,J.C.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - 20-45 DAP spike cDNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene SK primer.
FEATURES             Location/Qualifiers
     source          1. .392
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /cultivar="Chinese Spring"
                     /db_xref="taxon:4565"
                     /clone="WHE0859_E11_I21"
                     /tissue_type="Spike and seed"
                     /dev_stage="Adult plant"
                     /lab_host="E. coli SOLR"
                     /clone_lib="Wheat 20-45 DAP spike cDNA library"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; Plants were grown in the
                     greenhouse.  Spikes at 20 DAP and seeds at 30  to 45 DAP
                     were harvested, total RNA and poly(A) RNA were prepared, a
                     cDNA  library was made, and the cDNA clones were in vivo
                     excised to give pBluescript  phagemids in the TJ Close lab
                     (Choi, Close, Fenton) at the University of  California,
                     Riverside.  Plasmid DNA preparations and DNA sequencing
                     were  performed in the OD Anderson lab (all other
                     authors)."
ORIGIN

Alignment Scores:
Pred. No.:              2.23e+03        Length:         392
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-858-332G-28 (1-6) x BE637465 (1-392)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         78 CACAACCACAACCACAAT 95

RESULT 225
BW862216/c
LOCUS       BW862216                392 bp    mRNA    linear   EST 23-MAY-2005
DEFINITION  BW862216 Amphioxus Branchiostoma floridae unpublished cDNA library,
            neurula whole animal Branchiostoma floridae cDNA clone bbne066h19
            5', mRNA sequence.
ACCESSION   BW862216
VERSION     BW862216.1  GI:66470432
KEYWORDS    EST.
SOURCE      Branchiostoma floridae (Florida lancelet)
  ORGANISM  Branchiostoma floridae
            Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
            Branchiostoma.
REFERENCE   1  (bases 1 to 392)
  AUTHORS   Yu,J., Holland,L.Z., Shin-i,T., Kohara,Y., Satou,Y. and Satoh,N.
  TITLE     Expressed genes in Branchiostoma floridae
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .392
                     /organism="Branchiostoma floridae"
                     /mol_type="mRNA"
                     /db_xref="taxon:7739"
                     /clone="bbne066h19"
                     /tissue_type="whole animal"
                     /dev_stage="neurula"
                     /clone_lib="Amphioxus Branchiostoma floridae unpublished
                     cDNA library, neurula whole animal"
ORIGIN

Alignment Scores:
Pred. No.:              2.23e+03        Length:         392
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     5               Gaps:           0

US-09-858-332G-28 (1-6) x BW862216 (1-392)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        304 CATAATCATAATCATAAC 287

RESULT 226
AU040894/c
LOCUS       AU040894                393 bp    mRNA    linear   EST 04-DEC-1998
DEFINITION  AU040894 Mouse four-cell-embryo cDNA Mus musculus cDNA clone
            J0820C08 3', mRNA sequence.
ACCESSION   AU040894
VERSION     AU040894.1  GI:3954718
KEYWORDS    EST.
```

```
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 393)
  AUTHORS   Ko,M.S.H., Kitchen,J.R., Wang,X., Threat,T.A., Sun,T.,
            DePalma,G.E., Liang,Y., Kargul,G.J., Sharara,R., Lim,M.K. and
            Doi,H.
  TITLE     Systematic analyses of genes expressed in 4-cell mouse embryo (The
            ERATO/Doi Project at Wayne State University)
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Hirofumi Doi
            Doi Bioasymmetry Project, ERATO
            Japan Science and Technology Corporation (JST)
            WBG Marive East 12F, 2-6 Nakase, Mihama-ku, Chiba 261-71, Japan
            Email: hd@bioa.jst.go.jp.
FEATURES             Location/Qualifiers
     source          1. .393
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="J0820C08"
                     /dev_stage="four-cell-embryo"
                     /clone_lib="Mouse four-cell-embryo cDNA"
ORIGIN

Alignment Scores:
Pred. No.:                2.24e+03        Length:         393
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       1               Gaps:           0

US-09-858-332G-28 (1-6) x AU040894 (1-393)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        388 CACAACCACAACCACAAC 371

RESULT 227
AZ254386
LOCUS       AZ254386                393 bp    DNA     linear   GSS 31-JAN-2003
DEFINITION  Gm_ISb001_012_M06F ISU Soybean BAC Library (pBeloBAC11 HindIII)
            Glycine max genomic clone Glycine max genomic clone
            Gm_ISb001_012_M06, genomic survey sequence.
ACCESSION   AZ254386
VERSION     AZ254386.1  GI:8714460
KEYWORDS    GSS.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 393)
  AUTHORS   Larson,K., Mudge,J., Cooper,A., Grivna,S., Denny,R., Penuela,S.,
            Danesh,D. and Young,N.D.
  TITLE     BAC End sequences from a soybean genomic library (UMN)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Young Nevin D
            Department of Plant Pathology
            University of Minnesota
            495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul MN 55108, USA
            Tel: 612 625 2225
            Fax: 612 625 9728
            Email: neviny@tc.umn.edu
            Sequence on contig Gm_K636_ctg_a near RFLP K636_1 on linkage group
            A2. For more information, see Soybase at:
            http://soybase.agron.iastate.edu. Please see as an authority for
            the mapping/naming: Cregan P.B., T. Jarvik, A.L. Bush, R.C.
            Shoemaker, K.G. Lark, A.L. Kahler, N. Kaya, T.T. VanToai, D.G.
            Lohnes, J. Chung, and J.E. Specht. 1999a. An integrated genetic
            linkage map of the soybean genome. Crop Sci. 39:1464-1490
            Seq primer: M13F
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .393
                     /organism="Glycine max"
                     /mol_type="genomic DNA"
                     /cultivar="Williams82"
                     /db_xref="taxon:3847"
                     /clone="Gm_ISb001_012_M06"
                     /tissue_type="primary leaves"
                     /dev_stage="cotyledon"
                     /clone_lib="ISU Soybean BAC Library (pBeloBAC11 HindIII)
                     Glycine max genomic clone"
                     /note="Vector: pBeloBAC11; The ISU BAC library (Marek and
                     Shoemaker, Genome 40:420, 1997) was constructed using the
                     HindIII site in pBeloBAC11. The library consists of
                     approximately 40,000 clones with an average insert size of
                     150 Kb, equal to 4 haploid genome equivalents. Screening
                     of the library is done by hybridization of high-density
                     colony filters. Two high density filters, each containing
                     18,432 clones (doubly spotted) represent the portion of
                     the library screened"
ORIGIN

Alignment Scores:
Pred. No.:                2.24e+03        Length:         393
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       9               Gaps:           0

US-09-858-332G-28 (1-6) x AZ254386 (1-393)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         75 CATAACCATAACCATAAC 92

RESULT 228
AJ770749
LOCUS       AJ770749                394 bp    mRNA    linear   EST 06-AUG-2004
DEFINITION  AJ770749 Populus euphratica leaf 3-6 months, adult Populus
            euphratica cDNA clone P0001100016F02F1, mRNA sequence.
ACCESSION   AJ770749
VERSION     AJ770749.1  GI:50240733
KEYWORDS    EST.
SOURCE      Populus euphratica
  ORGANISM  Populus euphratica
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Malpighiales; Salicaceae; Saliceae; Populus.
REFERENCE   1  (bases 1 to 394)
  AUTHORS   Brosche,M., Alatalo,E.R., Vinocur,B., Altman,A., Teichmann,T.,
            Ottow,E.A., Polle,A., Djilianov,D., Afif,D., Triboulot,M.B.,
            Dreyer,E., Paulin,L. and Kangasjaervi,J.
  TITLE     Gene expression in desert leaves
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Prof. Jaakko Kangasjarvi
            Department of Biosciences, Plant Physiology
            University of Helsinki
            FIN-00014 Helsinki, Finland.
FEATURES             Location/Qualifiers
     source          1. .394
                     /organism="Populus euphratica"
                     /mol_type="mRNA"
                     /db_xref="taxon:75702"
                     /clone="P0001100016F02F1"
                     /tissue_type="leaf"
                     /dev_stage="3-6 months, adult"
                     /clone_lib="Populus euphratica leaf 3-6 months, adult"
```

```
                     /note="tissue_lib: P00011"
ORIGIN

Alignment Scores:
Pred. No.:                2.25e+03       Length:         394
Score:                    42.00          Matches:        6
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       1              Gaps:           0

US-09-858-332G-28 (1-6) x AJ770749 (1-394)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        153 CACAACCACAACCACAAT 170

RESULT 229
AU040954/c
LOCUS       AU040954                 394 bp    mRNA    linear   EST 04-DEC-1998
DEFINITION  AU040954 Mouse four-cell-embryo cDNA Mus musculus cDNA clone
            J0821E08 3', mRNA sequence.
ACCESSION   AU040954
VERSION     AU040954.1  GI:3955125
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 394)
  AUTHORS   Ko,M.S.H., Kitchen,J.R., Wang,X., Threat,T.A., Sun,T.,
            DePalma,G.E., Liang,Y., Kargul,G.J., Sharara,R., Lim,M.K. and
            Doi,H.
  TITLE     Systematic analyses of genes expressed in 4-cell mouse embryo (The
            ERATO/Doi Project at Wayne State University)
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Hirofumi Doi
            Doi Bioasymmetry Project, ERATO
            Japan Science and Technology Corporation (JST)
            WBG Marive East 12F, 2-6 Nakase, Mihama-ku, Chiba 261-71, Japan
            Email: hd@bioa.jst.go.jp.
FEATURES             Location/Qualifiers
     source          1. .394
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="J0821E08"
                     /dev_stage="four-cell-embryo"
                     /clone_lib="Mouse four-cell-embryo cDNA"
ORIGIN

Alignment Scores:
Pred. No.:                2.25e+03       Length:         394
Score:                    42.00          Matches:        6
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       1              Gaps:           0

US-09-858-332G-28 (1-6) x AU040954 (1-394)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        389 CACAACCACAACCACAAC 372

RESULT 230
AU040973/c
LOCUS       AU040973                 394 bp    mRNA    linear   EST 04-DEC-1998
DEFINITION  AU040973 Mouse four-cell-embryo cDNA Mus musculus cDNA clone
            J0821H03 3', mRNA sequence.
ACCESSION   AU040973
VERSION     AU040973.1  GI:3955144
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 394)
  AUTHORS   Ko,M.S.H., Kitchen,J.R., Wang,X., Threat,T.A., Sun,T.,
            DePalma,G.E., Liang,Y., Kargul,G.J., Sharara,R., Lim,M.K. and
            Doi,H.
  TITLE     Systematic analyses of genes expressed in 4-cell mouse embryo (The
            ERATO/Doi Project at Wayne State University)
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Hirofumi Doi
            Doi Bioasymmetry Project, ERATO
            Japan Science and Technology Corporation (JST)
            WBG Marive East 12F, 2-6 Nakase, Mihama-ku, Chiba 261-71, Japan
            Email: hd@bioa.jst.go.jp.
FEATURES             Location/Qualifiers
     source          1. .394
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="J0821H03"
                     /dev_stage="four-cell-embryo"
                     /clone_lib="Mouse four-cell-embryo cDNA"
ORIGIN

Alignment Scores:
Pred. No.:                2.25e+03       Length:         394
Score:                    42.00          Matches:        6
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       1              Gaps:           0

US-09-858-332G-28 (1-6) x AU040973 (1-394)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        389 CACAACCACAACCACAAC 372

RESULT 231
AU040007
LOCUS       AU040007                 395 bp    mRNA    linear   EST 29-MAR-1999
DEFINITION  AU040007 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
            discoideum cDNA clone SLH428, mRNA sequence.
ACCESSION   AU040007
VERSION     AU040007.1  GI:4009247
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 395)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     The Dictyostelium developmental cDNA project: generation and
            analysis of expressed sequence tags from the first-finger stage of
            development
  JOURNAL   DNA Res. 5 (6), 335-340 (1998)
   PUBMED   10048482
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp
            PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
```

```
     source          1. .395
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SLH428"
                     /dev_stage="slug"
                     /clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
ORIGIN

Alignment Scores:
Pred. No.:              2.25e+03        Length:         395
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AU040007 (1-395)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        267 CATAATCATAATCATAAT 284

RESULT 232
AU040972/c
LOCUS       AU040972                395 bp    mRNA    linear   EST 04-DEC-1998
DEFINITION  AU040972 Mouse four-cell-embryo cDNA Mus musculus cDNA clone
            J0821H02 3', mRNA sequence.
ACCESSION   AU040972
VERSION     AU040972.1  GI:3955143
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 395)
  AUTHORS   Ko,M.S.H., Kitchen,J.R., Wang,X., Threat,T.A., Sun,T.,
            DePalma,G.E., Liang,Y., Kargul,G.J., Sharara,R., Lim,M.K. and
            Doi,H.
  TITLE     Systematic analyses of genes expressed in 4-cell mouse embryo (The
            ERATO/Doi Project at Wayne State University)
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Hirofumi Doi
            Doi Bioasymmetry Project, ERATO
            Japan Science and Technology Corporation (JST)
            WBG Marive East 12F, 2-6 Nakase, Mihama-ku, Chiba 261-71, Japan
            Email: hd@bioa.jst.go.jp.
FEATURES             Location/Qualifiers
     source          1. .395
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="J0821H02"
                     /dev_stage="four-cell-embryo"
                     /clone_lib="Mouse four-cell-embryo cDNA"
ORIGIN

Alignment Scores:
Pred. No.:              2.25e+03        Length:         395
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AU040972 (1-395)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        389 CACAACCACAACCACAAC 372

RESULT 233
BP588507/c
LOCUS       BP588507                395 bp    mRNA    linear   EST 23-JUN-2004
DEFINITION  BP588507 RAFL15 Arabidopsis thaliana cDNA clone RAFL15-46-I02 3',
            mRNA sequence.
ACCESSION   BP588507
VERSION     BP588507.1  GI:49155975
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 395)
  AUTHORS   Seki,M., Narusaka,M., Kamiya,A., Ishida,J., Satou,M., Sakurai,T.,
            Nakajima,M., Enju,A., Akiyama,K., Oono,Y., Muramatsu,M.,
            Hayashizaki,Y., Kawai,J., Carninci,P., Itoh,M., Ishii,Y.,
            Arakawa,T., Shibata,K., Shinagawa,A. and Shinozaki,K.
  TITLE     Functional annotation of a full-length Arabidopsis cDNA collection
  JOURNAL   Science 296 (5565), 141-145 (2002)
  PUBMED    11910074
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            reversed clone; Please visit our web site
            (http://pfgweb.gsc.riken.go.jp/) for further details.
FEATURES             Location/Qualifiers
     source          1. .395
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="RAFL15-46-I02"
                     /tissue_type="mixture of silique and flower"
                     /lab_host="DH10B"
                     /clone_lib="RAFL15"
                     /note="Site_1: BamHI; Site_2: SalI"
ORIGIN

Alignment Scores:
Pred. No.:              2.25e+03        Length:         395
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-858-332G-28 (1-6) x BP588507 (1-395)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        305 CACAACCACAACCACAAC 288

RESULT 234
CF196074
LOCUS       CF196074                395 bp    mRNA    linear   EST 31-JUL-2003
DEFINITION  Azbi2.3_E03 AZ-BI Oryza sativa (japonica cultivar-group) cDNA clone
            AZ-BI2.3E03 5', mRNA sequence.
ACCESSION   CF196074
VERSION     CF196074.1  GI:33380731
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 395)
  AUTHORS   Ventelon-Debout,M., Thi-Thu Huang,N., Wissocq,A., Berger,C.,
            Michele,L., Piegu,B., Cooke,R., Ghesquiere,A., Delseny,M. and
```

```
            Brugidou,C.
  TITLE     Metabolism and Photosynthesis gene expression is affected in
            response to rice yellow mottle virus infection in Oryza sativa
            indica and japonica culltivars
  JOURNAL   Mol. Genet. Genomics (2003) In press
COMMENT     Contact: Brugidou C
            IRD
            BP 64501, 34394 Montpellier cedex 5 France
            Tel: 33 4 67 41 67 39
            Fax: 33 4 67 41 61 81
            PCR PRimers
            FORWARD: 5'AATTAACCCTCACTAAAGGG3'
            BACKWARD: 5'CGGGATATCACTCAGCATAATG3'
            Plate: 2.3  row: E  column: 3.
FEATURES             Location/Qualifiers
     source          1. .395
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Azucena"
                     /db_xref="taxon:39947"
                     /clone="AZ-BI2.3E03"
                     /tissue_type="leaves"
                     /dev_stage="3 weeks-old plants"
                     /lab_host="SOLR strain (Stratagene)"
                     /clone_lib="AZ-BI"
                     /note="Vector: pBluescript SK (+/-); Site_1: Eco RI;
                     Site_2: Xho I; Sample name: AZ-BI; Plant growth place:
                     greenhouse; Soil conditions: standard; Sowing date:
                     28/02/2000; Harvesting date: 18/03/2000; Stress date:
                     15/03/2000; Cultivar: azucena; Description: Az-wounded
                     leaves cDNA library; experimental condition: wound;"
ORIGIN

Alignment Scores:
Pred. No.:              2.25e+03        Length:         395
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x CF196074 (1-395)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         88 CACAACCACAACCACAAC 105

RESULT 235
BG139519
LOCUS       BG139519                396 bp    mRNA    linear   EST 31-JAN-2001
DEFINITION  EST480057 wild tomato pollen Lycopersicon pennellii cDNA clone
            cLPP13P21 5' sequence, mRNA sequence.
ACCESSION   BG139519
VERSION     BG139519.1  GI:12639803
KEYWORDS    EST.
SOURCE      Lycopersicon pennellii (Solanum pennellii)
  ORGANISM  Lycopersicon pennellii
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 396)
  AUTHORS   van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Utterback,T.,
            Hansen,C., Ronning,C. and Tanksley,S.
  TITLE     Generation of ESTs from wild tomato (L. pennellii) pollen
  JOURNAL   Unpublished (2001)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html.
FEATURES             Location/Qualifiers
     source          1. .396
                     /organism="Lycopersicon pennellii"
                     /mol_type="mRNA"
                     /cultivar="TA56"
                     /db_xref="taxon:28526"
                     /clone="cLPP13P21"
                     /tissue_type="pollen"
                     /dev_stage="pollen collected from open flowers"
                     /lab_host="SOLR"
                     /clone_lib="wild tomato pollen"
                     /note="Vector: pBluescript SK(-); Site_1: EcoR1; Site_2:
                     Xho1; Pollen was collected from open flowers from
                     L.pennellii TA56, and stored at -80 C until library
                     construction."
ORIGIN

Alignment Scores:
Pred. No.:              2.26e+03        Length:         396
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-858-332G-28 (1-6) x BG139519 (1-396)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        240 CACAACCACAACCACAAC 257

RESULT 236
BW876778/c
LOCUS       BW876778                396 bp    mRNA    linear   EST 24-MAY-2005
DEFINITION  BW876778 Amphioxus Branchiostoma floridae unpublished cDNA library,
            neurula whole animal Branchiostoma floridae cDNA clone bbne120107
            5', mRNA sequence.
ACCESSION   BW876778
VERSION     BW876778.1  GI:66491455
KEYWORDS    EST.
SOURCE      Branchiostoma floridae (Florida lancelet)
  ORGANISM  Branchiostoma floridae
            Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
            Branchiostoma.
REFERENCE   1  (bases 1 to 396)
  AUTHORS   Yu,J., Holland,L.Z., Shin-i,T., Kohara,Y., Satou,Y. and Satoh,N.
  TITLE     Expressed genes in Branchiostoma floridae
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .396
                     /organism="Branchiostoma floridae"
                     /mol_type="mRNA"
                     /db_xref="taxon:7739"
                     /clone="bbne120107"
                     /tissue_type="whole animal"
                     /dev_stage="neurula"
                     /clone_lib="Amphioxus Branchiostoma floridae unpublished
                     cDNA library, neurula whole animal"
ORIGIN

Alignment Scores:
Pred. No.:              2.26e+03        Length:         396
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     5               Gaps:           0

US-09-858-332G-28 (1-6) x BW876778 (1-396)
```

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        305 CATAATCATAATCATAAC 288

RESULT 237
CF959995
LOCUS       CF959995                396 bp    mRNA    linear   EST 15-JAN-2004
DEFINITION  8512rsiced_12146.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF959995
VERSION     CF959995.1  GI:38475862
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 396)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 396
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .396
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="trefoil"
                     /clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA
                     library"
ORIGIN

Alignment Scores:
Pred. No.:              2.26e+03        Length:         396
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     7               Gaps:           0

US-09-858-332G-28 (1-6) x CF959995 (1-396)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        199 CACAACCACAACCACAAC 216

RESULT 238
AI899397
LOCUS       AI899397                397 bp    mRNA    linear   EST 18-MAY-2001
DEFINITION  EST268840 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
            cLED38N10, mRNA sequence.
ACCESSION   AI899397
VERSION     AI899397.1  GI:5605299
KEYWORDS    EST.
SOURCE      Lycopersicon esculentum (Solanum lycopersicum)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 397)
  AUTHORS   Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T.,
            Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B.,
            Fujii,C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C.,
            Martin,G.B., Tanksley,S.D. and Giovannoni,J.
  TITLE     Generation of ESTs from tomato carpel tissue
  JOURNAL   Unpublished (1999)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html
            5 prime sequence.
FEATURES             Location/Qualifiers
     source          1. .397
                     /organism="Lycopersicon esculentum"
                     /mol_type="mRNA"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cLED38N10"
                     /tissue_type="carpel"
                     /dev_stage="5 days pre-anthesis to 5 days post-anthesis"
                     /lab_host="XL1-Blue MRF'"
                     /clone_lib="tomato ovary, TAMU"
                     /note="Vector: pBlueScript SK(-); Site_1: EcoR1; Site_2:
                     XhoI; cLED - Tomato Carpel EST Library. OligodT-primed and
                     directionally cloned  cDNA in vector Lamda ZAP II with 5'
                     and 3' ends located at the EcoRI and  XhoI sites,
                     respectively."
ORIGIN

Alignment Scores:
Pred. No.:              2.26e+03        Length:         397
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AI899397 (1-397)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         31 CACAACCACAATCACAAC 48

RESULT 239
CF966741
LOCUS       CF966741                397 bp    mRNA    linear   EST 15-JAN-2004
DEFINITION  12584rsiced_7618.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF966741
VERSION     CF966741.1  GI:38482939
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 397)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
```

```
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 397
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .397
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="trefoil"
                     /clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA
                     library"
ORIGIN

Alignment Scores:
Pred. No.:                2.26e+03        Length:         397
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       7               Gaps:           0

US-09-858-332G-28 (1-6) x CF966741 (1-397)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         75 CACAACCACAACCACAAC 92

RESULT 240
AA004078
LOCUS       AA004078                 398 bp    mRNA    linear   EST 23-JUL-1996
DEFINITION  mg81g02.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:439442 5' similar to gb:M38188 OVARIAN GRANULOSA CELL
            13.0 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION   AA004078
VERSION     AA004078.1  GI:1447553
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 398)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:264778
            Seq primer: mob.REGA+ET
            High quality sequence stop: 371.
FEATURES             Location/Qualifiers
     source          1. .398
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:439442"
                     /sex="unknown"
                     /tissue_type="embryo"
                     /dev_stage="13.5-14.5dpc total fetus"
                     /lab_host="DH10B"
                     /clone_lib="Soares mouse embryo NbME13.5 14.5"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTTT
                     T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                     14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                     State Univ., from 2 ]; double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector. Library went through one round of
                     normalization, and was constructed by Bento Soares and
                     M.Fatima Bonaldo.  "
ORIGIN

Alignment Scores:
Pred. No.:                2.27e+03        Length:         398
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       1               Gaps:           0

US-09-858-332G-28 (1-6) x AA004078 (1-398)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        315 CACAACCATAACCACAAC 332

RESULT 241
BY069921
LOCUS       BY069921                 398 bp    mRNA    linear   EST 06-DEC-2002
DEFINITION  BY069921 RIKEN full-length enriched, 17 days embryo kidney Mus
            musculus cDNA clone I920069N21 5', mRNA sequence.
ACCESSION   BY069921
VERSION     BY069921.1  GI:26172774
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 398)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
            Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
            Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
            Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
            Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
            Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
            Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
            Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
            Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
            Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
            Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
            Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
            Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
            Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
            Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
            Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
            Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
            Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
```

```
            Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
            Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
            Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
            Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
            Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
            Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
            Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
            Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
   PUBMED   12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
            Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
            Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,
            Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,
            Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and
            Hayashizaki,Y. Direct Submission
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
             cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES                Location/Qualifiers
     source             1. .398
                        /organism="Mus musculus"
                        /mol_type="mRNA"
                        /strain="C57BL/6J"
                        /db_xref="taxon:10090"
                        /clone="I920069N21"
                        /tissue_type="kidney"
                        /dev_stage="17 days embryo"
                        /clone_lib="RIKEN full-length enriched, 17 days embryo
                        kidney"
ORIGIN

Alignment Scores:
Pred. No.:                2.27e+03        Length:         398
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       5               Gaps:           0

US-09-858-332G-28 (1-6) x BY069921 (1-398)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        301 CACAACCATAACCACAAC 318

RESULT 242
AV428528
LOCUS       AV428528                399 bp    mRNA    linear   EST 10-MAR-2004
DEFINITION  AV428528 Lotus japonicus young plants (two-week old) Lotus
            corniculatus var. japonicus cDNA clone MWM098g03_r 5', mRNA
            sequence.
ACCESSION   AV428528
VERSION     AV428528.1  GI:7789573
KEYWORDS    EST.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1  (bases 1 to 399)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Generation of 7137 non-redundant expressed sequence tags from a
            legume, Lotus japonicus
  JOURNAL   DNA Res. 7 (2), 127-130 (2000)
   PUBMED   10819328
COMMENT     Contact: Yasukazu Nakamura
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: ynakamu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES                Location/Qualifiers
     source             1. .399
                        /organism="Lotus corniculatus var. japonicus"
                        /mol_type="mRNA"
                        /isolate="Miyakojima MG-20"
                        /db_xref="taxon:34305"
                        /clone="MWM098g03_r"
                        /tissue_type="whole plant"
                        /dev_stage="two-week old plant"
                        /clone_lib="Lotus japonicus young plants (two-week old)"
                        /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                        XhoI; synonym: Lotus japonicus"
ORIGIN

Alignment Scores:
Pred. No.:                2.28e+03        Length:         399
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       1               Gaps:           0

US-09-858-332G-28 (1-6) x AV428528 (1-399)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        272 CACAATCACAATCACAAT 289

RESULT 243
AL364520
LOCUS       AL364520                400 bp    mRNA    linear   EST 04-AUG-2000
DEFINITION  AL364520 ICRFp 522 and 523 Mus musculus cDNA clone B9C20D20 5',
            mRNA sequence.
ACCESSION   AL364520
VERSION     AL364520.1  GI:9691866
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 400)
  AUTHORS   Eickhoff,H., Schuchhardt,J., Ivanov,I., Meier-Ewert,S., O'Brien,J.,
            Malik,A., Tandon,N., Wolski,E., Rohlfs,E., Nyarsik,L.,
            Reinhardt,R., Nietfeld,W. and Lehrach,H.
  TITLE     Tissue gene expression analysis using arrayed normalized cDNA
            libraries
  JOURNAL   Genome Res. (2000) In press
COMMENT     Contact: MPIMG
```

```
           Abt.Lehrach
           Max Planck Institut Fuer Molekulare Genetik
           Ihnestrasse 73, Berlin, 14195 Germany
           The cDNA libraries ICRFp 522 and 523 were normalized with
           oligonucleotide fingerprinting, resulting in a unique subset of
           5376 cDNA clones.
FEATURES             Location/Qualifiers
     source          1. .400
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="B9C20D20"
                     /tissue_type="embryo"
                     /dev_stage="9 and 12 pc embryo"
                     /clone_lib="ICRFp 522 and 523"
ORIGIN

Alignment Scores:
Pred. No.:              2.28e+03        Length:         400
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AL364520 (1-400)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        311 CACAACCATAACCACAAC 328

RESULT 244
BF292873
LOCUS       BF292873                400 bp    mRNA    linear   EST 17-NOV-2000
DEFINITION  WHE2165_H07_P13ZS Triticum turgidum L. var. durum (durum wheat)
            whole plant cDNA library Triticum turgidum cDNA clone
            WHE2165_H07_P13, mRNA sequence.
ACCESSION   BF292873
VERSION     BF292873.1  GI:11223937
KEYWORDS    EST.
SOURCE      Triticum turgidum
  ORGANISM  Triticum turgidum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 400)
  AUTHORS   Akhunov,E., Anderson,O.D., Chao,S., Chin,A., Choi,D.W., Close,T.J.,
            Fenton,R.D., Han,P.S., Hsia,C.C., Kang,Y., Kianian,P., Lazo,G.R.,
            Miller,R., Otto,C., Rausch,C.J., Seaton,C.L., Simons,K., Tong,J.C.
            and Zhang,D.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Whole plant cDNA library from Triticum turgidum L. var.
            durum
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene SK primer.
FEATURES             Location/Qualifiers
     source          1. .400
                     /organism="Triticum turgidum"
                     /mol_type="mRNA"
                     /cultivar="Langdon-16"
                     /db_xref="taxon:4571"
                     /clone="WHE2165_H07_P13"
                     /tissue_type="All tissues"
                     /dev_stage="Different growth stages"
                     /lab_host="E. coli SOLR"
                     /clone_lib="Triticum turgidum L. var. durum (durum wheat)
                     whole plant cDNA library"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; Plants were grown in a growth
                     chamber at North Dakota State University  (Kianian, Otto,
                     Simons).  Tissues collected from seven-day etiolated
                     seedling leaf, stem, root and seed; leaf from plant at
                     fourth leaf  stage; spike from pre-anthesis through 20
                     days after anthesis; flag  leaf; leaf and stem tissue from
                     tillers. and root.  Total RNA and  poly(A) RNA were
                     prepared from each tissue and then pooled, a cDNA  library
                     was made, and the cDNA clones were in vivo excised to give
                     pBluescript phagemids in the TJ Close lab (Akhunov, Chin,
                     Choi, Close,  Fenton, Kianian, Otto, Simons, Zhang) at the
                     University of California,  Riverside.  Plasmid DNA
                     preparations and DNA sequencing were performed  in the OD
                     Anderson lab (all other authors)."
ORIGIN

Alignment Scores:
Pred. No.:              2.28e+03        Length:         400
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-858-332G-28 (1-6) x BF292873 (1-400)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        312 CACAACCACAACCACAAT 329

RESULT 245
BF293137
LOCUS       BF293137                400 bp    mRNA    linear   EST 17-NOV-2000
DEFINITION  WHE2164_A05_A10ZS Triticum turgidum L. var. durum (durum wheat)
            whole plant cDNA library Triticum turgidum cDNA clone
            WHE2164_A05_A10, mRNA sequence.
ACCESSION   BF293137
VERSION     BF293137.1  GI:11224201
KEYWORDS    EST.
SOURCE      Triticum turgidum
  ORGANISM  Triticum turgidum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 400)
  AUTHORS   Akhunov,E., Anderson,O.D., Chao,S., Chin,A., Choi,D.W., Close,T.J.,
            Fenton,R.D., Han,P.S., Hsia,C.C., Kang,Y., Kianian,P., Lazo,G.R.,
            Miller,R., Otto,C., Rausch,C.J., Seaton,C.L., Simons,K., Tong,J.C.
            and Zhang,D.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Whole plant cDNA library from Triticum turgidum L. var.
            durum
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene SK primer.
FEATURES             Location/Qualifiers
     source          1. .400
                     /organism="Triticum turgidum"
                     /mol_type="mRNA"
                     /cultivar="Langdon-16"
```

/db_xref="taxon:4571"
/clone="WHE2164_A05_A10"
/tissue_type="All tissues"
/dev_stage="Different growth stages"
/lab_host="E. coli SOLR"
/clone_lib="Triticum turgidum L. var. durum (durum wheat) whole plant cDNA library"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site_1: EcoRI; Site_2: XhoI; Plants were grown in a growth chamber at North Dakota State University (Kianian, Otto, Simons). Tissues collected from seven-day etiolated seedling leaf, stem, root and seed; leaf from plant at fourth leaf stage; spike from pre-anthesis through 20 days after anthesis; flag leaf; leaf and stem tissue from tillers. and root. Total RNA and poly(A) RNA were prepared from each tissue and then pooled, a cDNA library was made, and the cDNA clones were in vivo excised to give pBluescript phagemids in the TJ Close lab (Akhunov, Chin, Choi, Close, Fenton, Kianian, Otto, Simons, Zhang) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

ORIGIN

Alignment Scores:

Pred. No.:	2.28e+03	Length:	400
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-28 (1-6) x BF293137 (1-400)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         15 CACAACCACAACCACAAT 32
```

RESULT 246
CF542925
LOCUS CF542925 401 bp mRNA linear EST 22-SEP-2003
DEFINITION S014680-024-030-L14-SP6 MPIZ-ADIS-024-leaf Beta vulgaris cDNA clone 024-030-L14 5-PRIME, mRNA sequence.
ACCESSION CF542925
VERSION CF542925.1 GI:34891365
KEYWORDS EST.
SOURCE Beta vulgaris
ORGANISM Beta vulgaris
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; Caryophyllales; Amaranthaceae; Beta.
REFERENCE 1 (bases 1 to 401)
AUTHORS Herwig,R., Schulz,B., Weisshaar,B., Hennig,S., Steinfath,M., Drungowski,M., Stahl,D., Wruck,W., Menze,A., O'Brien,J., Lehrach,H. and Radelof,U.
TITLE Construction of a 'unigene' cDNA clone set by oligonucleotide fingerprinting allows access to 25 000 potential sugar beet genes
JOURNAL Plant J. 32 (5), 845-857 (2002)
PUBMED 12472698
COMMENT Contact: Weisshaar B
ADIS DNA core facility at MPIZ
Max-Planck-Institute for Plant Breeding Research
Carl-von-Linne Weg 10, 50829 Koeln, Germany
Fax: 00492215062851
Email: weisshaa@mpiz-koeln.mpg.de
Insert Length: 401 Std Error: 0.00
Plate: 30 row: L column: 14
Seq primer: SP6.
FEATURES Location/Qualifiers
source 1. .401
/organism="Beta vulgaris"
/mol_type="mRNA"
/cultivar="KWS2320 (double haploid, monogerm breeding line)"
/db_xref="GABI:936827"
/db_xref="taxon:161934"
/clone="024-030-L14"
/tissue_type="leaf"
/lab_host="EMDH10B"
/clone_lib="MPIZ-ADIS-024-leaf"
/note="Vector: pCMVSPORT6; Site_1: SalI; Site_2: NotI; cDNA library from sugar beet, library provided by KWS Kleinwanzlebener Saatzucht AG Einbeck, Germany, contact: b.schulz@kws.de; cloning sites SalI-NotI, primer sites and orientation:
SP6-SalI-CCACGCGTCCG-5prime-cDNA-polyA-CC-NotI-T7; Note: Sequencing granted in the context of the GABI-Beet project, local PI: Dr. Katharina Schneider, coordinator: Prof. Christian Jung; Sequence submission managed by RZPD/GABI-Primary database:http://gabi.rzpd.de"

ORIGIN

Alignment Scores:

Pred. No.:	2.29e+03	Length:	401
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-09-858-332G-28 (1-6) x CF542925 (1-401)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        110 CATAACCATAACCATAAC 127
```

RESULT 247
DR590716/c
LOCUS DR590716 401 bp mRNA linear EST 08-JUL-2005
DEFINITION WS00829.B21_F17 WS-X-N-A-9 Picea glauca cDNA clone WS00829_F17 3, mRNA sequence.
ACCESSION DR590716
VERSION DR590716.1 GI:70655459
KEYWORDS EST.
SOURCE Picea glauca (white spruce)
ORGANISM Picea glauca
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
REFERENCE 1 (bases 1 to 401)
AUTHORS Ralph,S., Kolosova,N., Oddy,C., Cooper,D., Butterfield,Y., Kirkpatrick,R., Liu,J., Palmquist,D., Stott,J., Barber,S., Yang,G., Babakaiff,R., Brown-John,M., Chand,S., Featherstone,R., Masson,A., Mayo,M., Moran,J., Olson,T., Wong,D., Friedmann,M.F., Ritland,C.E., Siddiqui,A., Holt,R., Jones,S., Marra,M., Ellis,B.E., Douglas,C., Ritland,K. and Bohlmann,J.
TITLE The spruce transcriptome: Analysis of expressed sequence tags from multiple cDNA libraries
JOURNAL Unpublished (2005)
COMMENT Contact: Joerg Bohlmann
Genome BC forest genomics program
University of British Columbia
Michael Smith Laboratories, 6174 University Boulevard, Rm. 237, Vancouver, British Columbia, Canada, V6T 1Z3
Tel: 1-604-822-0282
Fax: 1-604-822-2114
Email: bohlmann@msl.ubc.ca
Plate: WS00829 row: F column: 17
High quality sequence stop: 401
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .401
/organism="Picea glauca"
/mol_type="mRNA"
/cultivar="PG-29"
/db_xref="taxon:3330"
/clone="WS00829_F17"

/sex="Hermaphrodite"
/tissue_type="Early season xylem harvested June 15th, mid season xylem harvested July 10th and late season xylem harvested August 17th"
/lab_host="E. coli DH10B cells"
/clone_lib="WS-X-N-A-9"
/note="Organ: Outer xylem from 25 year old trees harvested at Kalamalka Research Station in Vernon, British Columbia in 2001; Vector: pBluescript II SK (+) XR; Site_1: EcoRI (5' end of cDNA); Site_2: XhoI (3' end of cDNA); mRNA was isolated from each tissue source independently and equal quantities of mRNA from each tissue were then pooled. cDNA was prepared from 5 micrograms of mRNA and directionally ligated into the pBluescript II SK (+) XR vector using the pBluescript II XR cDNA Library Construction Kit according to manufacturer's instructions with modifications (Stratagene). Plasmid DNA was then transformed by electroporation into DH10B cells (Invitrogen) for propagation. Normalization was applied according to published methods [Bonaldo M.F. et al. (1996) Genome Research 6(9):791] in order to reduce the abundance of highly expressed transcripts."

ORIGIN

Alignment Scores:

Pred. No.:	2.29e+03	Length:	401
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	8	Gaps:	0

US-09-858-332G-28 (1-6) x DR590716 (1-401)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        400 CACAATCACAATCACAAC 383
```

RESULT 248
AI408694
LOCUS AI408694 402 bp mRNA linear EST 09-FEB-1999
DEFINITION EST236985 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone ROVER75 3' end, mRNA sequence.
ACCESSION AI408694
VERSION AI408694.1 GI:4252198
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 402)
AUTHORS Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J., Kerlavage,A.R. and Adams,M.D.
TITLE Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat Gene Index
JOURNAL Unpublished (1998)
COMMENT Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
FEATURES Location/Qualifiers
source 1. .402
/organism="Rattus sp."
/mol_type="mRNA"
/db_xref="taxon:10118"
/clone="ROVER75"
/clone_lib="Normalized rat ovary, Bento Soares"
/note="Organ: ovary; Vector: pT7T3Pac; Site_1: EcoRI; Site_2: NotI"

ORIGIN

Alignment Scores:

Pred. No.:	2.29e+03	Length:	402
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0

US-09-858-332G-28 (1-6) x AI408694 (1-402)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        213 CATAATCACAATCATAAT 230
```

RESULT 249
BI623556
LOCUS BI623556 402 bp mRNA linear EST 07-SEP-2001
DEFINITION RH55856.5prime RH Drosophila melanogaster normalized Head pFlc-1 Drosophila melanogaster cDNA clone RH55856 5, mRNA sequence.
ACCESSION BI623556
VERSION BI623556.1 GI:15519081
KEYWORDS EST.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 402)
AUTHORS Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S., Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin,G.M.
TITLE BDGP/HHMI RH Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
hit genomic AE003473: arm:3L [1816241,2120346]
estimated-cyto:62B6-62C4: 08/22/2001
Plate: RH.558 row: E column: 8
High quality sequence stop: 339.
FEATURES Location/Qualifiers
source 1. .402
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="RH55856"
/sex="male and female"
/dev_stage="Adult"
/lab_host="DH5-alpha TonA"
/clone_lib="RH Drosophila melanogaster normalized Head pFlc-1"
/note="Organ: head; Vector: pFlc1; Site_1: XhoI; Site_2: BamHI; Library was kindly generated by Piero Carninci at the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."

ORIGIN

Alignment Scores:

Pred. No.:	2.29e+03	Length:	402
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	3	Gaps:	0

US-09-858-332G-28 (1-6) x BI623556 (1-402)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         33 CATAATCATAATCATAAT 50
```

RESULT 250
BM443064
LOCUS BM443064 402 bp mRNA linear EST 23-JUL-2002
DEFINITION EBro02_SQ002_G03_R root, 3 week, hydroponic grown, low nitrogen, cv Optic, EBro02 Hordeum vulgare subsp. vulgare cDNA clone EBro02_SQ002_G03 5', mRNA sequence.
ACCESSION BM443064
VERSION BM443064.2 GI:21947713
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 402)
AUTHORS Hedley,P., Liu,H., Caldwell,D., McCallum,N., Mudie,S., Cardle,L., Ramsay,L., Machray,G., Marshall,D.F.M. and Waugh,R.
TITLE Development of Barley Transcriptome Resources
JOURNAL Unpublished (2001)
COMMENT On Feb 1, 2002 this sequence version replaced gi:18473827.
Contact: Waugh R, Marshall DF
Genome Dynamics/Computational Biology
Scottish Crop Research Institute
Invergowrie, Dundee, DD2 5DA, Scotland, UK
Tel: 00 44 1382 562731
Fax: 00 44 1382 562426
Email: est@scri.sari.ac.uk
All sequence has a Phred quality score of 20 or over
Seq primer: M13 reverse.
FEATURES Location/Qualifiers
source 1. .402
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Optic"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="EBro02_SQ002_G03"
/tissue_type="root"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="root, 3 week, hydroponic grown, low nitrogen, cv Optic, EBro02"
/note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I; Non-normalised library, directionally cloned into pSPORT1. Derived from roots of 3 week old Nitrogen stressed barley plants. Developed as part of the barley transcriptome resources of BBSRC/SEERAD funded cereal IGF (Investigating Gene Function) project."
ORIGIN

Alignment Scores:

Pred. No.:	2.29e+03	Length:	402
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	3	Gaps:	0

US-09-858-332G-28 (1-6) x BM443064 (1-402)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         16 CACAATCACAATCACAAT 33
```

RESULT 251
AI490526
LOCUS AI490526 403 bp mRNA linear EST 18-MAY-2001
DEFINITION EST249060 tomato ovary, TAMU Lycopersicon esculentum cDNA clone cLED20N17, mRNA sequence.
ACCESSION AI490526
VERSION AI490526.1 GI:4385836
KEYWORDS EST.
SOURCE Lycopersicon esculentum (Solanum lycopersicum)
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 403)
AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
TITLE Generation of ESTs from tomato carpel tissue
JOURNAL Unpublished (1999)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html.
FEATURES Location/Qualifiers
source 1. .403
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLED20N17"
/tissue_type="carpel"
/dev_stage="5 days pre-anthesis to 5 days post-anthesis"
/lab_host="XL1-Blue MRF'"
/clone_lib="tomato ovary, TAMU"
/note="Vector: pBlueScript SK(-); Site_1: EcoR1; Site_2: Xho1; cLED - Tomato Carpel EST Library. OligodT-primed and directionally cloned cDNA in vector Lamda ZAP II with 5' and 3' ends located at the EcoRI and XhoI sites, respectively."
ORIGIN

Alignment Scores:

Pred. No.:	2.3e+03	Length:	403
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0

US-09-858-332G-28 (1-6) x AI490526 (1-403)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        260 CACAACCACAATCACAAC 277
```

RESULT 252
BG277971
LOCUS BG277971 403 bp mRNA linear EST 21-FEB-2001
DEFINITION ux64h12.y1 Soares_NKWMD_mandible Mus musculus cDNA clone IMAGE:3515374 5' similar to TR:Q9WTZ9 Q9WTZ9 BRAIN EXPRESSED X-LINKED PROTEIN 3. ;, mRNA sequence.
ACCESSION BG277971
VERSION BG277971.1 GI:13073809
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 403)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)

```
COMMENT     Other_ESTs: ux64h12.x1
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1395950
            Seq primer: -40RP from Gibco
            High quality sequence stop: 326.
FEATURES              Location/Qualifiers
     source           1. .403
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:3515374"
                      /tissue_type="mandible"
                      /lab_host="DH10B (phage-resistant)"
                      /clone_lib="Soares_NKWMD_mandible"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: NotI; Site_2: EcoRI; 1st strand cDNA
                      was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCCTTAATTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      went through one round of normalization. Library
                      constructed by Bento Soares and M. Fatima Bonaldo. "
ORIGIN

Alignment Scores:
Pred. No.:                2.3e+03        Length:          403
Score:                    42.00          Matches:         6
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       2              Gaps:            0

US-09-858-332G-28 (1-6) x BG277971 (1-403)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        296 CACAACCATAACCACAAC 313

RESULT 253
AQ121290
LOCUS       AQ121290                 403 bp    DNA     linear   GSS 22-SEP-1998
DEFINITION  HS_3075_A1_G09_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3075 Col=17 Row=M, genomic survey
            sequence.
ACCESSION   AQ121290
VERSION     AQ121290.1  GI:3498456
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 403)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3075  row: M  column: 17
            Class: BAC ends
            High quality sequence stop: 403.
FEATURES              Location/Qualifiers
     source           1. .403
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /clone="Plate=3075 Col=17 Row=M"
                      /sex="male"
                      /clone_lib="CIT Approved Human Genomic Sperm Library D"
                      /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                      E-Coli DH10B"
ORIGIN

Alignment Scores:
Pred. No.:                2.3e+03        Length:          403
Score:                    42.00          Matches:         6
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       9              Gaps:            0

US-09-858-332G-28 (1-6) x AQ121290 (1-403)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        280 CATAATCATAATCATAAT 297

RESULT 254
AA003697
LOCUS       AA003697                 404 bp    mRNA    linear   EST 22-JUL-1996
DEFINITION  mg60g09.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:437440 5' similar to gb:M38188 OVARIAN GRANULOSA CELL
            13.0 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION   AA003697
VERSION     AA003697.1  GI:1447153
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 404)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:262776
            Seq primer: ETPrimer
            High quality sequence stop: 355.
FEATURES              Location/Qualifiers
     source           1. .404
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:437440"
                      /sex="unknown"
                      /tissue_type="embryo"
                      /dev_stage="13.5-14.5dpc total fetus"
                      /lab_host="DH10B"
```

/clone_lib="Soares mouse embryo NbME13.5 14.5"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTTT T 3'], on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo. "
ORIGIN

Alignment Scores:

Pred. No.:	2.31e+03	Length:	404
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0

US-09-858-332G-28 (1-6) x AA003697 (1-404)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        259 CACAACCATAACCACAAC 276
```

RESULT 255
BP848836/c
LOCUS BP848836 404 bp mRNA linear EST 17-FEB-2005
DEFINITION BP848836 RAFL21 Arabidopsis thaliana cDNA clone RAFL25-06-N17 5', mRNA sequence.
ACCESSION BP848836
VERSION BP848836.1 GI:59909809
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 404)
AUTHORS Seki,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M., Akiyama,K., Iida,K., Enju,A., Sakurai,T., Arakawa,T., Carninci,P., Fukuda,S., Iida,J., Kawai,J., Sasaki,D., Shiraki,T., Hayashizaki,Y. and Shinozaki,K.
TITLE Large-Scale Analysis of RIKEN Arabidopsis Full-length cDNAs
JOURNAL Unpublished (2005)
COMMENT Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@rtc.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998, 2002).This clone is in a modified pBluescript vector.
Please visit our web site (http://pfgweb.gsc.riken.jp and http://rarge.gsc.riken.jp) for further details.
FEATURES Location/Qualifiers
source 1. .404
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAFL25-06-N17"
/lab_host="DH10B"
/clone_lib="RAFL21"
/note="Site_1: BamHI; Site_2: SalI; Subtraction Library. The sequence was obtained from samples subjected to various stress and plant hormones-treated"
ORIGIN

Alignment Scores:

Pred. No.:	2.31e+03	Length:	404
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	3	Gaps:	0

US-09-858-332G-28 (1-6) x BP848836 (1-404)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        144 CACAATCACAATCACAAT 127
```

RESULT 256
AI109877
LOCUS AI109877 405 bp mRNA linear EST 19-APR-2001
DEFINITION GH09260.5prime GH Drosophila melanogaster head pOT2 Drosophila melanogaster cDNA clone GH09260 5prime similar to X63731: D.melanogaster regulatory signal sequence (clone scs), mRNA sequence.
ACCESSION AI109877
VERSION AI109877.1 GI:3478201
KEYWORDS EST.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 405)
AUTHORS Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G., Lewis,S. and Rubin,G.M.
TITLE BDGP/HHMI Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 92 row: E column: 12
High quality sequence stop: 404.
FEATURES Location/Qualifiers
source 1. .405
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="GH09260"
/sex="male and female"
/dev_stage="adult"
/lab_host="DH5 - alpha"
/clone_lib="GH Drosophila melanogaster head pOT2"
/note="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2. Plasmid cDNA library."
ORIGIN

Alignment Scores:

Pred. No.:	2.31e+03	Length:	405
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0

US-09-858-332G-28 (1-6) x AI109877 (1-405)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        191 CACAACCATAATCACAAT 208
```

RESULT 257

```
BF201722
LOCUS       BF201722                405 bp    mRNA    linear   EST 06-NOV-2000
DEFINITION  WHE1774_H03_P06ZS Wheat pre-anthesis spike cDNA library Triticum
            aestivum cDNA clone WHE1774_H03_P06, mRNA sequence.
ACCESSION   BF201722
VERSION     BF201722.1  GI:11116464
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 405)
  AUTHORS   Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D.,
            Han,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J.,
            Seaton,C.L. and Tong,J.C.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Pre-anthesis spike cDNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene SK primer.
FEATURES             Location/Qualifiers
     source          1. .405
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /cultivar="Chinese Spring"
                     /db_xref="taxon:4565"
                     /clone="WHE1774_H03_P06"
                     /tissue_type="Spike before anthesis"
                     /dev_stage="Adult plant"
                     /lab_host="E. coli SOLR"
                     /clone_lib="Wheat pre-anthesis spike cDNA library"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; Plants were grown in the
                     greenhouse.  Whole spike with awns trimmed,  white, green
                     and yellow anther were collected and total RNA, and
                     poly(A) RNA were prepared, a cDNA library was made, and
                     the cDNA  clones were in vivo excised to give pBluescript
                     phagemids in the TJ  Close lab (Choi, Close, Fenton) at
                     the University of California,  Riverside.  Plasmid DNA
                     preparations and DNA sequencing were performed  in the OD
                     Anderson lab (all other authors)."
ORIGIN

Alignment Scores:
Pred. No.:              2.31e+03        Length:         405
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-858-332G-28 (1-6) x BF201722 (1-405)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         58 CACAACCACAACCACAAT 75

RESULT 258
BF324706
LOCUS       BF324706                405 bp    mRNA    linear   EST 13-JUL-2004
DEFINITION  su14c11.y1 Gm-c1066 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1066-1653 5' similar to TR:O65168 O65168 AUXIN-INDUCED PROTEIN
            ;, mRNA sequence.
ACCESSION   BF324706
VERSION     BF324706.1  GI:11274343
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 405)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,
            Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
            Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
            Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
            Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            When it has been determined, an EST from the other end of this
            clone is listed in the 'Other ESTs on clone' field. This clone is
            available through: Biogenetic Services, 801 32nd Ave. Brookings, SD
            57006 USA (phone: 800 423 4163; email: info@biogeneticservices.com)
            Insert Length: 745   Std Error: 0.00
            High quality sequence stop: 384.
FEATURES             Location/Qualifiers
     source          1. .405
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /cultivar="Williams"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1066-1653"
                     /tissue_type="Leaf and shoot tip, salt stressed, 2 week
                     old seedling"
                     /lab_host="DH10B"
                     /clone_lib="Gm-c1066"
                     /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library was constructed from mRNA isolated
                     from unexpanded leaves and the shoot tips of 2 week old
                     seedling from the cultivar Williams. The 2 week old
                     seedlings were salt stressed in a solution of 500mM NaCl
                     for 3 days prior to harvesting. Complementary DNA was
                     synthesized from mRNA using a  primer consisting of a
                     poly(dT) sequence with a XhoI restriction site. EcoRI
                     adapters were ligated to the blunt-ended cDNA fragments
                     followed by XhoI  digestion. The cDNA fragments were
                     directionally cloned into the EcoRI-XhoI  restriction site
                     of the pBluescript vector. The ligated cDNA fragments were
                     transformed into DH10B host cells (GibcoBRL). This library
                     was constructed  in the laboratory of Dr. Randy
                     Shoemaker."
ORIGIN

Alignment Scores:
Pred. No.:              2.31e+03        Length:         405
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-858-332G-28 (1-6) x BF324706 (1-405)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        298 CATAACCATAACCATAAC 315

RESULT 259
```

```
CD157052
LOCUS       CD157052                405 bp    mRNA    linear   EST 14-SEP-2003
DEFINITION  ML1-0050U-V132-A02-U.B ML1-0050 Schistosoma mansoni cDNA clone
            ML1-0050U-V132-A02.B, mRNA sequence.
ACCESSION   CD157052
VERSION     CD157052.1  GI:34693830
KEYWORDS    EST.
SOURCE      Schistosoma mansoni
  ORGANISM  Schistosoma mansoni
            Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
            Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE   1  (bases 1 to 405)
  AUTHORS   Verjovski-Almeida,S., DeMarco,R., Martins,E.A.L., Guimaraes,P.E.M.,
            Ojopi,E.P.B., Paquola,A.C.M., Piazza,J.P., Nishiyama,M.Y. Jr.,
            Kitajima,J.P., Adamson,R.E., Ashton,P.D., Bonaldo,M.F.,
            Coulson,P.S., Dillon,G.P., Farias,L.P., Gregorio,S.P., Ho,P.L.,
            Leite,R.A., Malaquias,L.C.C., Marques,R.C.P., Miyasato,P.A.,
            Nascimento,A.L.T.O., Ohlweiler,F.P., Reis,E.M., Ribeiro,M.A.,
            Sa,R.G., Stukart,G.C., Soares,M.B., Gargioni,C., Kawano,T.,
            Rodrigues,V., Madeira,A.M.B.N., Wilson,R.A., Menck,C.F.M.,
            Setubal,J.C., Leite,L.C.C. and Dias-Neto,E.
  TITLE     Transcriptome analysis of the acoelomate human parasite Schistosoma
            mansoni
  JOURNAL   Nat. Genet. 35 (2), 148-157 (2003)
   PUBMED   12973350
COMMENT     Contact: Dr. Sergio Verjovski-Almeida
            Departamento de Bioquimica
            Instituto de Quimica - Universidade de Sao Paulo
            Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 Sao Paulo - SP,
            Brasil
            Tel: +55-11-3091-2173
            Fax: +55-11-3091-2186
            Email: verjo@iq.usp.br
            This sequence was derived from the FAPESP Schistosoma mansoni EST
            Genome Project. All sequences in the project were assembled and
            annotated. This entry and all the assembled sequences can be seen
            in the following URL http://bioinfo.iq.usp.br/schisto/
            Plate: ML1-0050U-V132  row: 2  column: A.
FEATURES             Location/Qualifiers
     source          1. .405
                     /organism="Schistosoma mansoni"
                     /mol_type="mRNA"
                     /db_xref="taxon:6183"
                     /clone="ML1-0050U-V132-A02.B"
                     /sex="mixed pool"
                     /dev_stage="miracidium"
                     /clone_lib="ML1-0050"
                     /note="Vector: pGEM T-easy"
ORIGIN

Alignment Scores:
Pred. No.:                2.31e+03        Length:         405
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x CD157052 (1-405)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         50 CATAACCACAACCACAAT 67

RESULT 260
CD157053
LOCUS       CD157053                405 bp    mRNA    linear   EST 14-SEP-2003
DEFINITION  ML1-0050U-V132-A03-U.B ML1-0050 Schistosoma mansoni cDNA clone
            ML1-0050U-V132-A03.B, mRNA sequence.
ACCESSION   CD157053
VERSION     CD157053.1  GI:34693831
KEYWORDS    EST.
SOURCE      Schistosoma mansoni
  ORGANISM  Schistosoma mansoni
            Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
            Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE   1  (bases 1 to 405)
  AUTHORS   Verjovski-Almeida,S., DeMarco,R., Martins,E.A.L., Guimaraes,P.E.M.,
            Ojopi,E.P.B., Paquola,A.C.M., Piazza,J.P., Nishiyama,M.Y. Jr.,
            Kitajima,J.P., Adamson,R.E., Ashton,P.D., Bonaldo,M.F.,
            Coulson,P.S., Dillon,G.P., Farias,L.P., Gregorio,S.P., Ho,P.L.,
            Leite,R.A., Malaquias,L.C.C., Marques,R.C.P., Miyasato,P.A.,
            Nascimento,A.L.T.O., Ohlweiler,F.P., Reis,E.M., Ribeiro,M.A.,
            Sa,R.G., Stukart,G.C., Soares,M.B., Gargioni,C., Kawano,T.,
            Rodrigues,V., Madeira,A.M.B.N., Wilson,R.A., Menck,C.F.M.,
            Setubal,J.C., Leite,L.C.C. and Dias-Neto,E.
  TITLE     Transcriptome analysis of the acoelomate human parasite Schistosoma
            mansoni
  JOURNAL   Nat. Genet. 35 (2), 148-157 (2003)
   PUBMED   12973350
COMMENT     Contact: Dr. Sergio Verjovski-Almeida
            Departamento de Bioquimica
            Instituto de Quimica - Universidade de Sao Paulo
            Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 Sao Paulo - SP,
            Brasil
            Tel: +55-11-3091-2173
            Fax: +55-11-3091-2186
            Email: verjo@iq.usp.br
            This sequence was derived from the FAPESP Schistosoma mansoni EST
            Genome Project. All sequences in the project were assembled and
            annotated. This entry and all the assembled sequences can be seen
            in the following URL http://bioinfo.iq.usp.br/schisto/
            Plate: ML1-0050U-V132  row: 3  column: A.
FEATURES             Location/Qualifiers
     source          1. .405
                     /organism="Schistosoma mansoni"
                     /mol_type="mRNA"
                     /db_xref="taxon:6183"
                     /clone="ML1-0050U-V132-A03.B"
                     /sex="mixed pool"
                     /dev_stage="miracidium"
                     /clone_lib="ML1-0050"
                     /note="Vector: pGEM T-easy"
ORIGIN

Alignment Scores:
Pred. No.:                2.31e+03        Length:         405
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x CD157053 (1-405)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         50 CATAACCACAACCACAAT 67

RESULT 261
CD157059
LOCUS       CD157059                405 bp    mRNA    linear   EST 14-SEP-2003
DEFINITION  ML1-0050U-V132-A09-U.B ML1-0050 Schistosoma mansoni cDNA clone
            ML1-0050U-V132-A09.B, mRNA sequence.
ACCESSION   CD157059
VERSION     CD157059.1  GI:34693837
KEYWORDS    EST.
SOURCE      Schistosoma mansoni
  ORGANISM  Schistosoma mansoni
            Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
            Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE   1  (bases 1 to 405)
  AUTHORS   Verjovski-Almeida,S., DeMarco,R., Martins,E.A.L., Guimaraes,P.E.M.,
            Ojopi,E.P.B., Paquola,A.C.M., Piazza,J.P., Nishiyama,M.Y. Jr.,
            Kitajima,J.P., Adamson,R.E., Ashton,P.D., Bonaldo,M.F.,
            Coulson,P.S., Dillon,G.P., Farias,L.P., Gregorio,S.P., Ho,P.L.,
```

```
            Leite,R.A., Malaquias,L.C.C., Marques,R.C.P., Miyasato,P.A.,
            Nascimento,A.L.T.O., Ohlweiler,F.P., Reis,E.M., Ribeiro,M.A.,
            Sa,R.G., Stukart,G.C., Soares,M.B., Gargioni,C., Kawano,T.,
            Rodrigues,V., Madeira,A.M.B.N., Wilson,R.A., Menck,C.F.M.,
            Setubal,J.C., Leite,L.C.C. and Dias-Neto,E.
  TITLE     Transcriptome analysis of the acoelomate human parasite Schistosoma
            mansoni
  JOURNAL   Nat. Genet. 35 (2), 148-157 (2003)
   PUBMED   12973350
COMMENT     Contact: Dr. Sergio Verjovski-Almeida
            Departamento de Bioquimica
            Instituto de Quimica - Universidade de Sao Paulo
            Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 Sao Paulo - SP,
            Brasil
            Tel: +55-11-3091-2173
            Fax: +55-11-3091-2186
            Email: verjo@iq.usp.br
            This sequence was derived from the FAPESP Schistosoma mansoni EST
            Genome Project. All sequences in the project were assembled and
            annotated. This entry and all the assembled sequences can be seen
            in the following URL http://bioinfo.iq.usp.br/schisto/
            Plate: ML1-0050U-V132  row: 9  column: A.
FEATURES             Location/Qualifiers
     source          1. .405
                     /organism="Schistosoma mansoni"
                     /mol_type="mRNA"
                     /db_xref="taxon:6183"
                     /clone="ML1-0050U-V132-A09.B"
                     /sex="mixed pool"
                     /dev_stage="miracidium"
                     /clone_lib="ML1-0050"
                     /note="Vector: pGEM T-easy"
ORIGIN

Alignment Scores:
Pred. No.:                2.31e+03       Length:          405
Score:                    42.00          Matches:         6
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       6              Gaps:            0

US-09-858-332G-28 (1-6) x CD157059 (1-405)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         50 CATAACCACAACCACAAT 67

RESULT 262
CD157120
LOCUS       CD157120                405 bp    mRNA    linear   EST 14-SEP-2003
DEFINITION  ML1-0050U-V132-G10-U.B ML1-0050 Schistosoma mansoni cDNA clone
            ML1-0050U-V132-G10.B, mRNA sequence.
ACCESSION   CD157120
VERSION     CD157120.1  GI:34693897
KEYWORDS    EST.
SOURCE      Schistosoma mansoni
  ORGANISM  Schistosoma mansoni
            Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
            Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE   1  (bases 1 to 405)
  AUTHORS   Verjovski-Almeida,S., DeMarco,R., Martins,E.A.L., Guimaraes,P.E.M.,
            Ojopi,E.P.B., Paquola,A.C.M., Piazza,J.P., Nishiyama,M.Y. Jr.,
            Kitajima,J.P., Adamson,R.E., Ashton,P.D., Bonaldo,M.F.,
            Coulson,P.S., Dillon,G.P., Farias,L.P., Gregorio,S.P., Ho,P.L.,
            Leite,R.A., Malaquias,L.C.C., Marques,R.C.P., Miyasato,P.A.,
            Nascimento,A.L.T.O., Ohlweiler,F.P., Reis,E.M., Ribeiro,M.A.,
            Sa,R.G., Stukart,G.C., Soares,M.B., Gargioni,C., Kawano,T.,
            Rodrigues,V., Madeira,A.M.B.N., Wilson,R.A., Menck,C.F.M.,
            Setubal,J.C., Leite,L.C.C. and Dias-Neto,E.
  TITLE     Transcriptome analysis of the acoelomate human parasite Schistosoma
            mansoni
  JOURNAL   Nat. Genet. 35 (2), 148-157 (2003)
   PUBMED   12973350
COMMENT     Contact: Dr. Sergio Verjovski-Almeida
            Departamento de Bioquimica
            Instituto de Quimica - Universidade de Sao Paulo
            Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 Sao Paulo - SP,
            Brasil
            Tel: +55-11-3091-2173
            Fax: +55-11-3091-2186
            Email: verjo@iq.usp.br
            This sequence was derived from the FAPESP Schistosoma mansoni EST
            Genome Project. All sequences in the project were assembled and
            annotated. This entry and all the assembled sequences can be seen
            in the following URL http://bioinfo.iq.usp.br/schisto/
            Plate: ML1-0050U-V132  row: 10  column: G.
FEATURES             Location/Qualifiers
     source          1. .405
                     /organism="Schistosoma mansoni"
                     /mol_type="mRNA"
                     /db_xref="taxon:6183"
                     /clone="ML1-0050U-V132-G10.B"
                     /sex="mixed pool"
                     /dev_stage="miracidium"
                     /clone_lib="ML1-0050"
                     /note="Vector: pGEM T-easy"
ORIGIN

Alignment Scores:
Pred. No.:                2.31e+03       Length:          405
Score:                    42.00          Matches:         6
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       6              Gaps:            0

US-09-858-332G-28 (1-6) x CD157120 (1-405)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         50 CATAACCACAACCACAAT 67

RESULT 263
CD157232
LOCUS       CD157232                405 bp    mRNA    linear   EST 14-SEP-2003
DEFINITION  ML1-0050U-V160-D04-U.B ML1-0050 Schistosoma mansoni cDNA clone
            ML1-0050U-V160-D04.B, mRNA sequence.
ACCESSION   CD157232
VERSION     CD157232.1  GI:34694009
KEYWORDS    EST.
SOURCE      Schistosoma mansoni
  ORGANISM  Schistosoma mansoni
            Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
            Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE   1  (bases 1 to 405)
  AUTHORS   Verjovski-Almeida,S., DeMarco,R., Martins,E.A.L., Guimaraes,P.E.M.,
            Ojopi,E.P.B., Paquola,A.C.M., Piazza,J.P., Nishiyama,M.Y. Jr.,
            Kitajima,J.P., Adamson,R.E., Ashton,P.D., Bonaldo,M.F.,
            Coulson,P.S., Dillon,G.P., Farias,L.P., Gregorio,S.P., Ho,P.L.,
            Leite,R.A., Malaquias,L.C.C., Marques,R.C.P., Miyasato,P.A.,
            Nascimento,A.L.T.O., Ohlweiler,F.P., Reis,E.M., Ribeiro,M.A.,
            Sa,R.G., Stukart,G.C., Soares,M.B., Gargioni,C., Kawano,T.,
            Rodrigues,V., Madeira,A.M.B.N., Wilson,R.A., Menck,C.F.M.,
            Setubal,J.C., Leite,L.C.C. and Dias-Neto,E.
  TITLE     Transcriptome analysis of the acoelomate human parasite Schistosoma
            mansoni
  JOURNAL   Nat. Genet. 35 (2), 148-157 (2003)
   PUBMED   12973350
COMMENT     Contact: Dr. Sergio Verjovski-Almeida
            Departamento de Bioquimica
            Instituto de Quimica - Universidade de Sao Paulo
            Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 Sao Paulo - SP,
            Brasil
            Tel: +55-11-3091-2173
            Fax: +55-11-3091-2186
```

Email: verjo@iq.usp.br
This sequence was derived from the FAPESP Schistosoma mansoni EST Genome Project. All sequences in the project were assembled and annotated. This entry and all the assembled sequences can be seen in the following URL http://bioinfo.iq.usp.br/schisto/
Plate: ML1-0050U-V160 row: 4 column: D.

FEATURES Location/Qualifiers
source 1. .405
/organism="Schistosoma mansoni"
/mol_type="mRNA"
/db_xref="taxon:6183"
/clone="ML1-0050U-V160-D04.B"
/sex="mixed pool"
/dev_stage="miracidium"
/clone_lib="ML1-0050"
/note="Vector: pGEM T-easy"

ORIGIN

Alignment Scores:

Pred. No.:	2.31e+03	Length:	405
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-09-858-332G-28 (1-6) x CD157232 (1-405)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         50 CATAACCACAACCACAAT 67
```

RESULT 264
CD157273
LOCUS CD157273 405 bp mRNA linear EST 14-SEP-2003
DEFINITION ML1-0050U-V160-H11-U.B ML1-0050 Schistosoma mansoni cDNA clone ML1-0050U-V160-H11.B, mRNA sequence.
ACCESSION CD157273
VERSION CD157273.1 GI:34694050
KEYWORDS EST.
SOURCE Schistosoma mansoni
ORGANISM Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE 1 (bases 1 to 405)
AUTHORS Verjovski-Almeida,S., DeMarco,R., Martins,E.A.L., Guimaraes,P.E.M., Ojopi,E.P.B., Paquola,A.C.M., Piazza,J.P., Nishiyama,M.Y. Jr., Kitajima,J.P., Adamson,R.E., Ashton,P.D., Bonaldo,M.F., Coulson,P.S., Dillon,G.P., Farias,L.P., Gregorio,S.P., Ho,P.L., Leite,R.A., Malaquias,L.C.C., Marques,R.C.P., Miyasato,P.A., Nascimento,A.L.T.O., Ohlweiler,F.P., Reis,E.M., Ribeiro,M.A., Sa,R.G., Stukart,G.C., Soares,M.B., Gargioni,C., Kawano,T., Rodrigues,V., Madeira,A.M.B.N., Wilson,R.A., Menck,C.F.M., Setubal,J.C., Leite,L.C.C. and Dias-Neto,E.
TITLE Transcriptome analysis of the acoelomate human parasite Schistosoma mansoni
JOURNAL Nat. Genet. 35 (2), 148-157 (2003)
PUBMED 12973350
COMMENT Contact: Dr. Sergio Verjovski-Almeida
Departamento de Bioquimica
Instituto de Quimica - Universidade de Sao Paulo
Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 Sao Paulo - SP, Brasil
Tel: +55-11-3091-2173
Fax: +55-11-3091-2186
Email: verjo@iq.usp.br
This sequence was derived from the FAPESP Schistosoma mansoni EST Genome Project. All sequences in the project were assembled and annotated. This entry and all the assembled sequences can be seen in the following URL http://bioinfo.iq.usp.br/schisto/
Plate: ML1-0050U-V160 row: 11 column: H.

FEATURES Location/Qualifiers
source 1. .405
/organism="Schistosoma mansoni"
/mol_type="mRNA"
/db_xref="taxon:6183"
/clone="ML1-0050U-V160-H11.B"
/sex="mixed pool"
/dev_stage="miracidium"
/clone_lib="ML1-0050"
/note="Vector: pGEM T-easy"

ORIGIN

Alignment Scores:

Pred. No.:	2.31e+03	Length:	405
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-09-858-332G-28 (1-6) x CD157273 (1-405)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         50 CATAACCACAACCACAAT 67
```

RESULT 265
CD157278
LOCUS CD157278 405 bp mRNA linear EST 14-SEP-2003
DEFINITION ML1-0050U-V182-A05-U.B ML1-0050 Schistosoma mansoni cDNA clone ML1-0050U-V182-A05.B, mRNA sequence.
ACCESSION CD157278
VERSION CD157278.1 GI:34694055
KEYWORDS EST.
SOURCE Schistosoma mansoni
ORGANISM Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE 1 (bases 1 to 405)
AUTHORS Verjovski-Almeida,S., DeMarco,R., Martins,E.A.L., Guimaraes,P.E.M., Ojopi,E.P.B., Paquola,A.C.M., Piazza,J.P., Nishiyama,M.Y. Jr., Kitajima,J.P., Adamson,R.E., Ashton,P.D., Bonaldo,M.F., Coulson,P.S., Dillon,G.P., Farias,L.P., Gregorio,S.P., Ho,P.L., Leite,R.A., Malaquias,L.C.C., Marques,R.C.P., Miyasato,P.A., Nascimento,A.L.T.O., Ohlweiler,F.P., Reis,E.M., Ribeiro,M.A., Sa,R.G., Stukart,G.C., Soares,M.B., Gargioni,C., Kawano,T., Rodrigues,V., Madeira,A.M.B.N., Wilson,R.A., Menck,C.F.M., Setubal,J.C., Leite,L.C.C. and Dias-Neto,E.
TITLE Transcriptome analysis of the acoelomate human parasite Schistosoma mansoni
JOURNAL Nat. Genet. 35 (2), 148-157 (2003)
PUBMED 12973350
COMMENT Contact: Dr. Sergio Verjovski-Almeida
Departamento de Bioquimica
Instituto de Quimica - Universidade de Sao Paulo
Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 Sao Paulo - SP, Brasil
Tel: +55-11-3091-2173
Fax: +55-11-3091-2186
Email: verjo@iq.usp.br
This sequence was derived from the FAPESP Schistosoma mansoni EST Genome Project. All sequences in the project were assembled and annotated. This entry and all the assembled sequences can be seen in the following URL http://bioinfo.iq.usp.br/schisto/
Plate: ML1-0050U-V182 row: 5 column: A.

FEATURES Location/Qualifiers
source 1. .405
/organism="Schistosoma mansoni"
/mol_type="mRNA"
/db_xref="taxon:6183"
/clone="ML1-0050U-V182-A05.B"
/sex="mixed pool"
/dev_stage="miracidium"
/clone_lib="ML1-0050"
/note="Vector: pGEM T-easy"

```
ORIGIN

Alignment Scores:
Pred. No.:              2.31e+03        Length:         405
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x CD157278 (1-405)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         50 CATAACCACAACCACAAT 67

RESULT 266
CF195899
LOCUS       CF195899                405 bp    mRNA    linear   EST 31-JUL-2003
DEFINITION  Azbi2.1_A12 AZ-BI Oryza sativa (japonica cultivar-group) cDNA clone
            AZ-BI2.1A12 5', mRNA sequence.
ACCESSION   CF195899
VERSION     CF195899.1  GI:33380556
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 405)
  AUTHORS   Ventelon-Debout,M., Thi-Thu Huang,N., Wissocq,A., Berger,C.,
            Michele,L., Piegu,B., Cooke,R., Ghesquiere,A., Delseny,M. and
            Brugidou,C.
  TITLE     Metabolism and Photosynthesis gene expression is affected in
            response to rice yellow mottle virus infection in Oryza sativa
            indica and japonica culltivars
  JOURNAL   Mol. Genet. Genomics (2003) In press
COMMENT     Contact: Brugidou C
            IRD
            BP 64501, 34394 Montpellier cedex 5 France
            Tel: 33 4 67 41 67 39
            Fax: 33 4 67 41 61 81
            PCR PRimers
            FORWARD: 5'AATTAACCCTCACTAAAGGG3'
            BACKWARD: 5'CGGGATATCACTCAGCATAATG3'
            Plate: 2.1  row: A  column: 12.
FEATURES             Location/Qualifiers
     source          1. .405
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Azucena"
                     /db_xref="taxon:39947"
                     /clone="AZ-BI2.1A12"
                     /tissue_type="leaves"
                     /dev_stage="3 weeks-old plants"
                     /lab_host="SOLR strain (Stratagene)"
                     /clone_lib="AZ-BI"
                     /note="Vector: pBluescript SK (+/-); Site_1: Eco RI;
                     Site_2: Xho I; Sample name: AZ-BI; Plant growth place:
                     greenhouse; Soil conditions: standard; Sowing date:
                     28/02/2000; Harvesting date: 18/03/2000; Stress date:
                     15/03/2000; Cultivar: azucena; Description: Az-wounded
                     leaves cDNA library; experimental condition: wound;"
ORIGIN

Alignment Scores:
Pred. No.:              2.31e+03        Length:         405
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x CF195899 (1-405)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         53 CACAACCACAACCACAAC 70

RESULT 267
CW693980/c
LOCUS       CW693980                405 bp    DNA     linear   GSS 03-NOV-2004
DEFINITION  OG_BBa0053N17.r OG_BBa Oryza glaberrima genomic clone OG_BBa0053N17
            3', genomic survey sequence.
ACCESSION   CW693980
VERSION     CW693980.1  GI:55277858
KEYWORDS    GSS.
SOURCE      Oryza glaberrima (African rice)
  ORGANISM  Oryza glaberrima
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 405)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Byrne,M., Stum,D., Smart,D., Rao,K.,
            Luo,M., Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C.
            and Wing,R.
  TITLE     OMAP
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: rwing@genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0053  row: N  column: 17
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .405
                     /organism="Oryza glaberrima"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4538"
                     /clone="OG_BBa0053N17"
                     /tissue_type="young leaves"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OG_BBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

Alignment Scores:
Pred. No.:              2.31e+03        Length:         405
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     10              Gaps:           0

US-09-858-332G-28 (1-6) x CW693980 (1-405)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         56 CACAACCACAACCACAAT 39

RESULT 268
BJ428470/c
LOCUS       BJ428470                406 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ428470 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv12f02 3', mRNA sequence.
ACCESSION   BJ428470
VERSION     BJ428470.1  GI:19403192
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
```

```
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 406)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .406
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv12f02"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:                2.32e+03        Length:         406
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       3               Gaps:           0

US-09-858-332G-28 (1-6) x BJ428470 (1-406)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        332 CACAATCACAATCACAAT 315

RESULT 269
B23897/c
LOCUS       B23897                   406 bp    DNA     linear   GSS 10-OCT-1997
DEFINITION  F17L16TR IGF Arabidopsis thaliana genomic clone F17L16, genomic
            survey sequence.
ACCESSION   B23897
VERSION     B23897.1  GI:2509528
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 406)
  AUTHORS   Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Coffey,E.,
            Golden,K., Johnson,K., Adams,M.D. and Venter,J.C.
  TITLE     A BAC End Sequence Database for Identifying Minimal Overlaps in
            Arabidopsis Genomic Sequencing
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: F17L16TF
            Contact: Steve Rounsley
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: rounsley@tigr.org
            Seq primer: M13 Reverse
            Class: BAC ends
            High quality sequence stop: 406.
FEATURES             Location/Qualifiers
     source          1. .406
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /ecotype="Columbia"
                     /db_xref="taxon:3702"
                     /clone="F17L16"
                     /sex="hermaphrodite"
                     /clone_lib="IGF"
                     /note="Vector: BeloBACII; Site_1: EcoRI; Site_2: EcoRI;
                     Produced by Thomas Altmann"
ORIGIN

Alignment Scores:
Pred. No.:                2.32e+03        Length:         406
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       9               Gaps:           0

US-09-858-332G-28 (1-6) x B23897 (1-406)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        258 CATAACCATAACCATAAT 241

RESULT 270
AV811060/c
LOCUS       AV811060                 408 bp    mRNA    linear   EST 29-MAR-2002
DEFINITION  AV811060 RAFL9 Arabidopsis thaliana cDNA clone RAFL09-66-I06 3',
            mRNA sequence.
ACCESSION   AV811060
VERSION     AV811060.1  GI:19845045
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 408)
  AUTHORS   Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
            Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,
            Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y.
            and Shinozaki,K.
  TITLE     Large scale analysis of Arabidopsis full-length cDNA (2002b)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            An Arabidopsis full-length cDNA library was constructed essentially
            as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
            and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et
            al., submitted for publication) digested with BamHI and SalI. This
            clone is in a modified pBluescript vector. Please visit our web
            site (http://www.gsc.riken.go.jp/e/plant/index _e.html) for further
            details.
FEATURES             Location/Qualifiers
     source          1. .408
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="RAFL09-66-I06"
                     /dev_stage="plants at various developmental stages from
                     germination to mature seeds"
                     /lab_host="DH10B"
                     /clone_lib="RAFL9"
                     /note="Site_1: BamHI; Site_2: SalI; subjected to
                     dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24
                     hr) treatments"
ORIGIN
```

```
Alignment Scores:
Pred. No.:              2.33e+03        Length:         408
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AV811060 (1-408)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        293 CACAACCACAACCACAAC 276

RESULT 271
BP028270
LOCUS       BP028270                408 bp    mRNA    linear   EST 19-AUG-2004
DEFINITION  BP028270 Lotus corniculatus var. japonicus flower Lotus
            corniculatus var. japonicus cDNA clone MF001d10_f 3', mRNA
            sequence.
ACCESSION   BP028270
VERSION     BP028270.1  GI:45405430
KEYWORDS    EST.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1  (bases 1 to 408)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Characteristics of the Lotus japonicus gene repertoire deduced from
            large-scale expressed sequence tag (EST) analysis
  JOURNAL   Plant Mol. Biol. 54 (3), 405-414 (2004)
   PUBMED   15284495
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .408
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="mRNA"
                     /isolate="Miyakojima MG-20"
                     /db_xref="taxon:34305"
                     /clone="MF001d10_f"
                     /tissue_type="flower"
                     /clone_lib="Lotus corniculatus var. japonicus flower"
ORIGIN

Alignment Scores:
Pred. No.:              2.33e+03        Length:         408
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-858-332G-28 (1-6) x BP028270 (1-408)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        288 CATAATCATAATCATAAT 305

RESULT 272
BQ843997
LOCUS       BQ843997                408 bp    mRNA    linear   EST 14-AUG-2002
DEFINITION  QGA12E19.yg.ab1 QG_ABCDI lettuce salinas Lactuca sativa cDNA clone
            QGA12E19, mRNA sequence.
ACCESSION   BQ843997
VERSION     BQ843997.1  GI:22224326
KEYWORDS    EST.
SOURCE      Lactuca sativa
  ORGANISM  Lactuca sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
            Cichorieae; Lactuca.
REFERENCE   1  (bases 1 to 408)
  AUTHORS   Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J.,
            Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y.,
            Lai,Z., Church,S., Jackson,L. and Bradford,K.
  TITLE     Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
            belongs to contig QG_CA_Contig3158, see http://cgpdb.ucdavis.edu/
            for details.
            Plate: QGA12  row: E  column: 19.
FEATURES             Location/Qualifiers
     source          1. .408
                     /organism="Lactuca sativa"
                     /mol_type="mRNA"
                     /cultivar="Salinas"
                     /db_xref="taxon:4236"
                     /clone="QGA12E19"
                     /lab_host="E.coli"
                     /clone_lib="QG_ABCDI lettuce salinas"
                     /note="Vector: pBRcDNASfiAB; The library was constructed
                     from 10 different sources of RNA from a single genotype.
                     Separate cDNAs were generated using primers that
                     incorporated unique 5' and 3' tags to distinguish each
                     source of RNA. cDNAs were then pooled, size-fractionated,
                     directionally cloned into a custom medium-copy vector and
                     transformations made with four size classes to minimize
                     size bias. Details of each source of RNA and library
                     construction can be obtained at http://cgpdb.ucdavis.edu/
                     TAG_TISSUE=roots
                     TAG_LIB=QG_ABCDI lettuce salinas
                     TAG_SEQ=GTTGCACGGG"
ORIGIN

Alignment Scores:
Pred. No.:              2.33e+03        Length:         408
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     5               Gaps:           0

US-09-858-332G-28 (1-6) x BQ843997 (1-408)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         58 CACAACCACAACCACAAC 75

RESULT 273
BP049169
LOCUS       BP049169                409 bp    mRNA    linear   EST 19-AUG-2004
DEFINITION  BP049169 Lotus corniculatus var. japonicus pods (less than 20 mm in
            length) Lotus corniculatus var. japonicus cDNA clone SPD065c03_f
            3', mRNA sequence.
ACCESSION   BP049169
VERSION     BP049169.1  GI:45605849
KEYWORDS    EST.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
```

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae; Lotus.
REFERENCE 1 (bases 1 to 409)
AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
TITLE Characteristics of the Lotus japonicus gene repertoire deduced from large-scale expressed sequence tag (EST) analysis
JOURNAL Plant Mol. Biol. 54 (3), 405-414 (2004)
PUBMED 15284495
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES Location/Qualifiers
source 1. .409
/organism="Lotus corniculatus var. japonicus"
/mol_type="mRNA"
/isolate="Miyakojima MG-20"
/db_xref="taxon:34305"
/clone="SPD065c03_f"
/tissue_type="pods"
/dev_stage="pod (less than 20 mm in length)"
/clone_lib="Lotus corniculatus var. japonicus pods (less than 20 mm in length)"
ORIGIN

Alignment Scores:

Pred. No.:	2.34e+03	Length:	409
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	3	Gaps:	0

US-09-858-332G-28 (1-6) x BP049169 (1-409)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db          4 CATAATCACAATCATAAT 21
```

RESULT 274
AA475682
LOCUS AA475682 410 bp mRNA linear EST 18-JUN-1997
DEFINITION vh16b04.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone IMAGE:875599 5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION AA475682
VERSION AA475682.1 GI:2203533
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 410)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:515079
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 152.
FEATURES Location/Qualifiers
source 1. .410
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:875599"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/clone_lib="Soares_mammary_gland_NbMMG"
/note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."
ORIGIN

Alignment Scores:

Pred. No.:	2.34e+03	Length:	410
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0

US-09-858-332G-28 (1-6) x AA475682 (1-410)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        154 CACAACCATAACCACAAC 171
```

RESULT 275
BP029304
LOCUS BP029304 410 bp mRNA linear EST 19-AUG-2004
DEFINITION BP029304 Lotus corniculatus var. japonicus flower Lotus corniculatus var. japonicus cDNA clone MF020c02_f 3', mRNA sequence.
ACCESSION BP029304
VERSION BP029304.1 GI:45406464
KEYWORDS EST.
SOURCE Lotus corniculatus var. japonicus (Lotus japonicus)
ORGANISM Lotus corniculatus var. japonicus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae; Lotus.
REFERENCE 1 (bases 1 to 410)
AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
TITLE Characteristics of the Lotus japonicus gene repertoire deduced from large-scale expressed sequence tag (EST) analysis
JOURNAL Plant Mol. Biol. 54 (3), 405-414 (2004)
PUBMED 15284495
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES Location/Qualifiers
source 1. .410
/organism="Lotus corniculatus var. japonicus"
/mol_type="mRNA"
/isolate="Miyakojima MG-20"
/db_xref="taxon:34305"

```
                     /clone="MF020c02_f"
                     /tissue_type="flower"
                     /clone_lib="Lotus corniculatus var. japonicus flower"
ORIGIN

Alignment Scores:
Pred. No.:                2.34e+03        Length:         410
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       3               Gaps:           0

US-09-858-332G-28 (1-6) x BP029304 (1-410)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        290 CATAATCATAATCATAAT 307

RESULT 276
BY288035
LOCUS       BY288035                411 bp    mRNA    linear   EST 11-DEC-2002
DEFINITION  BY288035 RIKEN full-length enriched, visual cortex Mus musculus
            cDNA clone K530044F14 5', mRNA sequence.
ACCESSION   BY288035
VERSION     BY288035.1  GI:26478372
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 411)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
            Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
            Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
            Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
            Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
            Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
            Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
            Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
            Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
            Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
            Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
            Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
            Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
            Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
            Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
            Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
            Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
            Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
            Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
            Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
            Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
            Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
            Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
            Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
            Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
            Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
   PUBMED   12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
            Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
            Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,
            Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,
            Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and
            Hayashizaki,Y. Direct Submission
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Tissues were provided by Michela Fagiolini and Takao K. Hensch (
            Laboratory for Neuronal Circuit Development Brain Science Institute
            RIKEN 2-1 Hirosawa,Wako-shi,Saitama 351-0198 Japan ) whose
            assistance we gratefully acknowledge. Please visit our web site
            (http://genome.gsc.riken.go.jp) for further details.
FEATURES              Location/Qualifiers
     source           1. .411
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="K530044F14"
                      /tissue_type="visual cortex"
                      /clone_lib="RIKEN full-length enriched, visual cortex"
ORIGIN

Alignment Scores:
Pred. No.:                2.35e+03        Length:         411
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       5               Gaps:           0

US-09-858-332G-28 (1-6) x BY288035 (1-411)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        304 CACAACCATAACCACAAC 321

RESULT 277
AQ010728/c
LOCUS       AQ010728                411 bp    DNA     linear   GSS 29-MAY-1998
DEFINITION  F27P2TFB IGF Arabidopsis thaliana genomic clone F27P2, genomic
            survey sequence.
ACCESSION   AQ010728
VERSION     AQ010728.1  GI:3166005
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 411)
  AUTHORS   Rounsley,S.D., Suh,E.J., Wible,C., Golden,K., Shatsman,S., Choi,P.,
            Yu,K., Akinretoye,B., Shen,K., Goonasekaram,S., Militscher,J.,
            Adams,M.D. and Venter,J.C.
  TITLE     A BAC End Sequence Database for Identifying Minimal Overlaps in
            Arabidopsis Genomic Sequencing. Update 4
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: F27P2TF F27P2TR
            Contact: Steve Rounsley
```

```
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: rounsley@tigr.org
            Seq primer: M13-21
            Class: BAC ends
            High quality sequence stop: 411.
FEATURES             Location/Qualifiers
     source          1. .411
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /ecotype="Columbia"
                     /db_xref="taxon:3702"
                     /clone="F27P2"
                     /sex="hermaphrodite"
                     /clone_lib="IGF"
                     /note="Vector: BeloBACII; Site_1: EcoRI; Site_2: EcoRI;
                     Produced by Thomas Altmann"
ORIGIN

Alignment Scores:
Pred. No.:                2.35e+03        Length:         411
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       9               Gaps:           0

US-09-858-332G-28 (1-6) x AQ010728 (1-411)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        255 CATAACCATAACCATAAT 238

RESULT 278
FR0042428/c
LOCUS       FR0042428               411 bp    DNA     linear   GSS 25-FEB-2004
DEFINITION  Fugu rubripes GSS sequence, clone 151J19yA3, genomic survey
            sequence.
ACCESSION   AL129920
VERSION     AL129920.1  GI:6111866
KEYWORDS    GSS; genome survey sequence.
SOURCE      Takifugu rubripes (Fugu rubripes)
  ORGANISM  Takifugu rubripes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Takifugu.
REFERENCE   1
  AUTHORS   Elgar,G., Clark,M.S., Meek,S., Smith,S., Warner,S., Edwards,Y.J.,
            Bouchireb,N., Cottage,A., Yeo,G.S., Umrania,Y., Williams,G. and
            Brenner,S.
  TITLE     Generation and analysis of 25 Mb of genomic DNA from the pufferfish
            Fugu rubripes by sequence scanning
  JOURNAL   Genome Res. 9 (10), 960-971 (1999)
   PUBMED   10523524
REFERENCE   2  (bases 1 to 411)
  AUTHORS   Elgar,G., Clark,M.S., Smith,S., Meek,S., Warner,S., Edwards,Y.J.K.,
            Umrania,Y., Williams,G. and Brenner,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource
            Centre, Hinxton, Cambridge, CB10 1SB. UK Email:
            biohelp@hgmp.mrc.ac.uk
COMMENT     Vector: pBluescript II KS
            V_type: phagemid
            PRIMER: KS
            DESCR:
            One pass dye-terminator sequencing of cosmid cloned genomic
            sequence.
FEATURES             Location/Qualifiers
     source          1. .411
                     /organism="Takifugu rubripes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:31033"
                     /clone="151J19yA3"
                     /clone_lib="cosmid 151J19"
ORIGIN

Alignment Scores:
Pred. No.:                2.35e+03        Length:         411
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       11              Gaps:           0

US-09-858-332G-28 (1-6) x FR0042428 (1-411)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        118 CATAACCATAACCATAAC 101

RESULT 279
W81787
LOCUS       W81787                  412 bp    mRNA    linear   EST 12-SEP-1996
DEFINITION  me92f05.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:403041 5' similar to gb:M38188 OVARIAN GRANULOSA CELL
            13.0 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION   W81787
VERSION     W81787.1  GI:1539256
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 412)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1392806.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:246809
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 405.
FEATURES             Location/Qualifiers
     source          1. .412
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:403041"
                     /sex="unknown"
                     /tissue_type="embryo"
                     /dev_stage="13.5-14.5dpc total fetus"
                     /lab_host="DH10B"
                     /clone_lib="Soares mouse embryo NbME13.5 14.5"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTTTT
                     T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
```

14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo. "

ORIGIN

Alignment Scores:

Pred. No.:	2.36e+03	Length:	412
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	8	Gaps:	0

US-09-858-332G-28 (1-6) x W81787 (1-412)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         48 CACAACCATAACCACAAC 65
```

RESULT 280
FR0042427
LOCUS FR0042427 412 bp DNA linear GSS 25-FEB-2004
DEFINITION Fugu rubripes GSS sequence, clone 151J19yA4, genomic survey sequence.
ACCESSION AL129919
VERSION AL129919.1 GI:6111865
KEYWORDS GSS; genome survey sequence.
SOURCE Takifugu rubripes (Fugu rubripes)
ORGANISM Takifugu rubripes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetradontoidea; Tetraodontidae; Takifugu.
REFERENCE 1
AUTHORS Elgar,G., Clark,M.S., Meek,S., Smith,S., Warner,S., Edwards,Y.J., Bouchireb,N., Cottage,A., Yeo,G.S., Umrania,Y., Williams,G. and Brenner,S.
TITLE Generation and analysis of 25 Mb of genomic DNA from the pufferfish Fugu rubripes by sequence scanning
JOURNAL Genome Res. 9 (10), 960-971 (1999)
PUBMED 10523524
REFERENCE 2 (bases 1 to 412)
AUTHORS Elgar,G., Clark,M.S., Smith,S., Meek,S., Warner,S., Edwards,Y.J.K., Umrania,Y., Williams,G. and Brenner,S.
TITLE Direct Submission
JOURNAL Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource Centre, Hinxton, Cambridge, CB10 1SB. UK Email: biohelp@hgmp.mrc.ac.uk
COMMENT Vector: pBluescript II KS
V_type: phagemid
PRIMER: KS
DESCR:
One pass dye-terminator sequencing of cosmid cloned genomic sequence.
FEATURES Location/Qualifiers
source 1. .412
/organism="Takifugu rubripes"
/mol_type="genomic DNA"
/db_xref="taxon:31033"
/clone="151J19yA4"
/clone_lib="cosmid 151J19"
ORIGIN

Alignment Scores:

Pred. No.:	2.36e+03	Length:	412
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	11	Gaps:	0

US-09-858-332G-28 (1-6) x FR0042427 (1-412)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        273 CATAACCATAACCATAAC 290
```

RESULT 281
AV788333/c
LOCUS AV788333 413 bp mRNA linear EST 28-MAR-2002
DEFINITION AV788333 RAFL6 Arabidopsis thaliana cDNA clone RAFL06-79-B02 3', mRNA sequence.
ACCESSION AV788333
VERSION AV788333.1 GI:19807123
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 413)
AUTHORS Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M., Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y., Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y. and Shinozaki,K.
TITLE Large scale analysis of Arabidopsis full-length cDNA (2002b)
JOURNAL Unpublished (2002)
COMMENT Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@rtc.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998). This clone is in a modified pBluescript vector as a SstI/XhoI insert. Please visit our web site (http://www.gsc.riken.go.jp/e/plant/index _e.html) for further details.
FEATURES Location/Qualifiers
source 1. .413
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAFL06-79-B02"
/dev_stage="plants at various developmental stages from germination to mature seeds"
/lab_host="DH10B"
/clone_lib="RAFL6"
/note="Site_1: SstI; Site_2: XhoI; subjected to dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24 hr) treatments"
ORIGIN

Alignment Scores:

Pred. No.:	2.36e+03	Length:	413
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0

US-09-858-332G-28 (1-6) x AV788333 (1-413)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        297 CACAACCACAACCACAAC 280
```

RESULT 282
BG669342
LOCUS BG669342 413 bp mRNA linear EST 30-APR-2001
DEFINITION DRNAAED03 Rat DRG Library Rattus norvegicus cDNA clone DRNAAED03

```
            5', mRNA sequence.
ACCESSION   BG669342
VERSION     BG669342.1  GI:13891264
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   Xiao,H.S., Huang,Q.H., Zhang,F.X., Bao,L., Lu,Y.J., Guo,C.,
            Yang,L., Huang,W.J., Fu,G., Xu,S.H., Cheng,X.P., Yan,Q., Zhu,Z.D.,
            Zhang,X., Chen,Z., Han,Z.G. and Zhang,X.
  TITLE     Identification of gene expression profile of dorsal root ganglion
            in the rat peripheral axotomy model of neuropathic pain
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (12), 8360-8366 (2002)
   PUBMED   12060780
COMMENT     Contact: Zhang Xu
            Laboratory of Sensory System
            Institute of Neuroscience
            320 Yue Yang Road, Shanghai 200031, P.R.China
            Tel: 86-21-64748700-121
            Fax: 86-21-64713446
            Email: xu.zhang@ion.ac.cn
            This clone is also available at Chinese National Human Genome
            Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park,
            Pudong New Area, P.R.China. Please contact with Zhang Xu
            (xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)
            PCR PRimers
            FORWARD: T3
            BACKWARD: T7
            Seq primer: T3
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .413
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="DRNAAED03"
                     /sex="male"
                     /tissue_type="dorsal root ganglion"
                     /dev_stage="adult"
                     /clone_lib="Rat DRG Library"
ORIGIN

Alignment Scores:
Pred. No.:              2.36e+03        Length:         413
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-858-332G-28 (1-6) x BG669342 (1-413)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        213 CACAACCATAACCACAAC 230

RESULT 283
CN193652/c
LOCUS       CN193652                413 bp    mRNA    linear   EST 05-APR-2004
DEFINITION  rg90a05.y1 Meloidogyne hapla female SMART pGEM Meloidogyne hapla
            cDNA 5' similar to contains element MER30 repetitive element ;,
            mRNA sequence.
ACCESSION   CN193652
VERSION     CN193652.1  GI:46218591
KEYWORDS    EST.
SOURCE      Meloidogyne hapla
  ORGANISM  Meloidogyne hapla
            Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
            Tylenchoidea; Meloidogynidae; Meloidogyninae; Meloidogyne.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J.,
            Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B.,
            Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C.,
            Tsagareishvili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C.,
            Underwood,K., Steptoe,M., Allen,M., Person,B., Swaller,T.,
            Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     The Washington Univ. Nematode EST Project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: McCarter JP
            The Washington Univ. Nematode EST Project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Cloned unidirectionally. Poly(A)+ RNA was concentrated and purified
            using Dynabeads (Dynal) and mRNA eluted for first strand synthesis.
            First strand cDNA was created using MMLV RT (Powerscript, Clontech)
            and primed with oligo(dT) with XhoI site and 5'SMART 'anchor' added
            using chimeric DNA-RNA oligo. 12 PCR cycles were done using first
            strand and primers specific to SMART oligo and 3' end. Double
            stranded cDNA was digested using XhoI/NotI, fractioned on
            Chroma-spin 400 columns (Clontech) and ligated to digested
            pGEM-11zf(+) plasmid. Chemically competent DH10B cells were used as
            host cells. Library materials provided by Dr. David Bird of North
            Carolina State University. Library construction by Jeff Rousch. See
            www.nematode.net for additional project information.
            Putative full length read
            The vector to vector length is 433
            Seq primer: Sp6.
FEATURES             Location/Qualifiers
     source          1. .413
                     /organism="Meloidogyne hapla"
                     /mol_type="mRNA"
                     /db_xref="taxon:6305"
                     /tissue_type="whole organism"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="Meloidogyne hapla female SMART pGEM"
                     /note="Vector: plasmid (ampicillin resistant); Site_1:
                     XhoI; Site_2: NotI; Cloned unidirectionally. Poly(A)+ RNA
                     was concentrated and purified using Dynabeads (Dynal) and
                     mRNA eluted for first strand synthesis. First strand cDNA
                     was created using MMLV RT (Powerscript, Clontech) and
                     primed with oligo(dT) with XhoI site and 5'SMART 'anchor'
                     added using chimeric DNA-RNA oligo. 12 PCR cycles were
                     done using first strand and primers specific to SMART
                     oligo and 3' end. Double stranded cDNA was digested using
                     XhoI/NotI, fractioned on Chroma-spin 400 columns
                     (Clontech) and ligated to digested pGEM-11zf(+) plasmid.
                     Chemically competent DH10B cells were used as host cells.
                     Library materials provided by Dr. David Bird of North
                     Carolina State University. Library construction by Jeff
                     Rousch. See www.nematode.net for additional project
                     information."
ORIGIN

Alignment Scores:
Pred. No.:              2.36e+03        Length:         413
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     7               Gaps:           0

US-09-858-332G-28 (1-6) x CN193652 (1-413)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        257 CATAATCACAATCATAAT 240
```

```
RESULT 284
AQ121286
LOCUS       AQ121286                413 bp    DNA     linear   GSS 22-SEP-1998
DEFINITION  HS_3075_A1_G04_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3075 Col=7 Row=M, genomic survey
            sequence.
ACCESSION   AQ121286
VERSION     AQ121286.1  GI:3498452
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3075  row: M  column: 7
            Class: BAC ends
            High quality sequence stop: 413.
FEATURES             Location/Qualifiers
     source          1. .413
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=3075 Col=7 Row=M"
                     /sex="male"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
ORIGIN

Alignment Scores:
Pred. No.:              2.36e+03        Length:         413
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     9               Gaps:           0

US-09-858-332G-28 (1-6) x AQ121286 (1-413)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        287 CATAATCATAATCATAAT 304

RESULT 285
BQ864704
LOCUS       BQ864704                414 bp    mRNA    linear   EST 14-AUG-2002
DEFINITION  QGC27H01.yg.ab1 QG_ABCDI lettuce salinas Lactuca sativa cDNA clone
            QGC27H01, mRNA sequence.
ACCESSION   BQ864704
VERSION     BQ864704.1  GI:22250169
KEYWORDS    EST.
SOURCE      Lactuca sativa
  ORGANISM  Lactuca sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
            Cichorieae; Lactuca.
REFERENCE   1  (bases 1 to 414)
  AUTHORS   Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J.,
            Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y.,
            Lai,Z., Church,S., Jackson,L. and Bradford,K.
  TITLE     Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
            belongs to contig QG_CA_Contig6998, see http://cgpdb.ucdavis.edu/
            for details.
            Plate: QGC27  row: H  column: 01.
FEATURES             Location/Qualifiers
     source          1. .414
                     /organism="Lactuca sativa"
                     /mol_type="mRNA"
                     /cultivar="Salinas"
                     /db_xref="taxon:4236"
                     /clone="QGC27H01"
                     /lab_host="E.coli"
                     /clone_lib="QG_ABCDI lettuce salinas"
                     /note="Vector: pBRcDNASfiAB; The library was constructed
                     from 10 different sources of RNA from a single genotype.
                     Separate cDNAs were generated using primers that
                     incorporated unique 5' and 3' tags to distinguish each
                     source of RNA. cDNAs were then pooled, size-fractionated,
                     directionally cloned into a custom medium-copy vector and
                     transformations made with four size classes to minimize
                     size bias. Details of each source of RNA and library
                     construction can be obtained at http://cgpdb.ucdavis.edu/
                     TAG_TISSUE=chemical induction
                     TAG_LIB=QG_ABCDI lettuce salinas
                     TAG_SEQ=TGTAGCCGGG"
ORIGIN

Alignment Scores:
Pred. No.:              2.37e+03        Length:         414
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     5               Gaps:           0

US-09-858-332G-28 (1-6) x BQ864704 (1-414)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        378 CATAATCATAATCATAAT 395

RESULT 286
BX002883
LOCUS       BX002883                415 bp    DNA     linear   GSS 02-APR-2004
DEFINITION  Arabidopsis thaliana T-DNA flanking sequence GK-367G12-017041,
            genomic survey sequence.
ACCESSION   BX002883
VERSION     BX002883.1  GI:26187843
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Li,Y., Rosso,M.G., Strizhov,N., Viehoever,P. and Weisshaar,B.
  TITLE     GABI-Kat SimpleSearch: a flanking sequence tag (FST) database for
            the identification of T-DNA insertion mutants in Arabidopsis
            thaliana
```

```
  JOURNAL   Bioinformatics 19 (11), 1441-1442 (2003)
   PUBMED   12874060
REFERENCE   2
  AUTHORS   Rosso,M.G., Li,Y., Strizhov,N., Reiss,B., Dekker,K. and
            Weisshaar,B.
  TITLE     An Arabidopsis thaliana T-DNA mutagenized population (GABI-Kat) for
            flanking sequence tag-based reverse genetics
  JOURNAL   Plant Mol. Biol. 53 (1-2), 247-259 (2003)
   PUBMED   14756321
REFERENCE   3
  AUTHORS   Strizhov,N., Li,Y., Rosso,M.G., Viehoever,P., Dekker,K.A. and
            Weisshaar,B.
  TITLE     High-throughput generation of sequence indexes from T-DNA
            mutagenized Arabidopsis thaliana lines
  JOURNAL   BioTechniques 35 (6), 1164-1168 (2003)
   PUBMED   14682050
REFERENCE   4  (bases 1 to 415)
  AUTHORS   Strizhov,N., Li,Y., Rosso,M.G. and Weisshaar,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-2004) Weisshaar B., Max-Planck-Institut fuer
            Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
COMMENT     This sequence has been recovered from the left border of the T-DNA.
            It indicates an insertion within the locus defined by BAC clone
            f17h15. Details on the protocols used for generation of the
            sequence are described in References 1-3. The sequences are
            generated at the MPI for Plant Breeding Research in the context of
            the GABI-Kat project. GABI-Kat is part of the German Plant Genomics
            program designated 'GABI'. Information on line availability can be
            found at: http://www.mpiz-koeln.mpg.de/GABI-Kat/.
FEATURES             Location/Qualifiers
     source          1. .415
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /clone="GK-367G12-017041"
                     /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
                     /ecotype="Col-0"
                     /note="PCR was performed on DNA from Arabidopsis thaliana
                     plants (T1) which were transformed with the T-DNA from
                     vector pAC161 (GenBank accession number: AJ537514). The
                     lines contain one or more T-DNA insertions. The DNA
                     fragment(s) resulting from the PCR were directly sequenced
                     to determine the genomic sequence flanking the insertion.
                     T-DNA derived sequences were removed."
ORIGIN

Alignment Scores:
Pred. No.:                2.37e+03        Length:          415
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       10              Gaps:            0

US-09-858-332G-28 (1-6) x BX002883 (1-415)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        130 CATAATCATAACCATAAC 147

RESULT 287
BM270929
LOCUS       BM270929                 416 bp    mRNA    linear   EST 06-JUL-2004
DEFINITION  sak03c05.y1 Gm-c1074 Glycine max cDNA clone SOYBEAN CLONE ID:
            Gm-c1074-5097 5' similar to TR:O65168 O65168 AUXIN-INDUCED PROTEIN
            ;, mRNA sequence.
ACCESSION   BM270929
VERSION     BM270929.1  GI:17964190
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 416)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,
            Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
            Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
            Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
            Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            When it has been determined, an EST from the other end of this
            clone is listed in the 'Other ESTs on clone' field. This clone is
            available through: Biogenetic Services, 801 32nd Ave. Brookings, SD
            57006 USA (phone: 800 423 4163; email: info@biogeneticservices.com)
            Putative full length read vector to vector length is 477
            Seq primer: -40RP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .416
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /cultivar="Williams 82"
                     /db_xref="taxon:3847"
                     /clone="SOYBEAN CLONE ID: Gm-c1074-5097"
                     /tissue_type="seedlings induced for HR (hypersensitive
                     response)"
                     /dev_stage="9-11 day old"
                     /lab_host="DH10B"
                     /clone_lib="Gm-c1074"
                     /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library was constructed from mRNA isolated
                     from 9-11 day old seedlings that were induced for HR
                     (hypersensitive response) by vacuum infiltrating plant
                     tissue with Pseudomonas syringae pv. glycinea carrying the
                     avrB gene (Genetics 141:1597-1604). Plant tissue (expanded
                     unifoliate leaves) was collected at 2, 4, 8, 12, 24, 36,
                     and 53 hrs after inoculation and their mRNA pooled equally
                     for cDNA construction. The  library was prepared using the
                     Stratagene pBluescript II SK(+) library construction kit.
                     Complementary DNA was synthesized from mRNA using a primer
                     consisting of a poly(dT) sequence with an XhoI restriction
                     site. EcoRI adaptors were ligated to the blunt-ended cDNA
                     fragments followed by XhoI digestion. The cDNA insert is
                     protected from XhoI digestion via methylation during first
                     strand synthesis. The cDNA fragments were directionally
                     cloned into the EcoRI-XhoI restriction site of the
                     pBluescript vector. The ligated cDNA fragments were
                     transformed into E.coli ElectroMax DH10B host cells. Plant
                     care, inoculations, and library construction were
                     performed by Steve Clough (Lila Vodkin lab, University of
                     Illinois)."
ORIGIN

Alignment Scores:
Pred. No.:                2.38e+03        Length:          416
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       3               Gaps:            0

US-09-858-332G-28 (1-6) x BM270929 (1-416)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        240 CATAACCATAACCATAAC 257
```

```
RESULT 288
BY247437
LOCUS       BY247437                416 bp    mRNA    linear   EST 10-DEC-2002
DEFINITION  BY247437 RIKEN full-length enriched, visual cortex Mus musculus
            cDNA clone K230310K01 5', mRNA sequence.
ACCESSION   BY247437
VERSION     BY247437.1  GI:26428949
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 416)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
            Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
            Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
            Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
            Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
            Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
            Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
            Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
            Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
            Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
            Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
            Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
            Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
            Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
            Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
            Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
            Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
            Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
            Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
            Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
            Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
            Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
            Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
            Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
            Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
            Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
   PUBMED   12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
            Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
            Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,
            Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,
            Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and
            Hayashizaki,Y. Direct Submission
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
             cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
             Tissues were provided by Michela Fagiolini and Takao K. Hensch (
            Laboratory for Neuronal Circuit Development Brain Science Institute
            RIKEN 2-1 Hirosawa,Wako-shi,Saitama 351-0198 Japan ) whose
            assistance we gratefully acknowledge. Please visit our web site
            (http://genome.gsc.riken.go.jp) for further details.
FEATURES            Location/Qualifiers
     source         1. .416
                    /organism="Mus musculus"
                    /mol_type="mRNA"
                    /strain="C57BL/6J"
                    /db_xref="taxon:10090"
                    /clone="K230310K01"
                    /tissue_type="visual cortex"
                    /clone_lib="RIKEN full-length enriched, visual cortex"
ORIGIN

Alignment Scores:
Pred. No.:              2.38e+03        Length:         416
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     5               Gaps:           0

US-09-858-332G-28 (1-6) x BY247437 (1-416)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        327 CACAACCATAACCACAAC 344

RESULT 289
AU034405
LOCUS       AU034405                417 bp    mRNA    linear   EST 28-APR-1999
DEFINITION  AU034405 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
            discoideum cDNA clone SLC829, mRNA sequence.
ACCESSION   AU034405
VERSION     AU034405.1  GI:3799829
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 417)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     The Dictyostelium developmental cDNA project: generation and
            analysis of expressed sequence tags from the first-finger stage of
            development
  JOURNAL   DNA Res. 5 (6), 335-340 (1998)
   PUBMED   10048482
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp
            PROJECT = 'Dictyostelium discoideum cDNA project in Japan'
            POLYA=No.
FEATURES            Location/Qualifiers
     source         1. .417
                    /organism="Dictyostelium discoideum"
                    /mol_type="mRNA"
                    /strain="AX4"
                    /db_xref="taxon:44689"
                    /clone="SLC829"
                    /dev_stage="slug"
                    /clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
ORIGIN
```

```
Alignment Scores:
Pred. No.:                2.39e+03        Length:         417
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       1               Gaps:           0

US-09-858-332G-28 (1-6) x AU034405 (1-417)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        129 CACAATCACAATCACAAT 146

RESULT 290
AW159517
LOCUS       AW159517                417 bp    mRNA    linear   EST 05-NOV-1999
DEFINITION  zb03d07.x50 Xenopus EST library Xenopus laevis cDNA clone zb03d07
            5', mRNA sequence.
ACCESSION   AW159517
VERSION     AW159517.1  GI:6271546
KEYWORDS    EST.
SOURCE      Xenopus laevis (African clawed frog)
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Xenopus.
REFERENCE   1  (bases 1 to 417)
  AUTHORS   Schutz,K., de la Bastide,M., Huang,E.N., Nascimento,L., Preston,R.,
            Shah,R., Swaby,I., Shekher,M., Spiegel,L., Vil,M.D. and
            McCombie,W.R.
  TITLE     Expressed sequence tags from Xenopus
  JOURNAL   Unpublished (1999)
COMMENT     Contact: W. Richard McCombie
            Lita Annenberg Hazen Genome Sequencing Center
            Cold Spring Harbor Laboratory
            PO Box 100, Cold Spring Harbor, NY 11724, USA
            Tel: 516 367 8884
            Fax: 516 367 8874
            Email: mccombie@cshl.org
            Plate: zb03  row: d  column: 07
            Seq primer: M13 universal forward primer
            High quality sequence stop: 417.
FEATURES             Location/Qualifiers
     source          1. .417
                     /organism="Xenopus laevis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8355"
                     /clone="zb03d07"
                     /tissue_type="total brain tissue"
                     /cell_line="W22-TGA"
                     /dev_stage="tadpole"
                     /clone_lib="Xenopus EST library"
                     /note="Vector: Lambda Zap I; Site_1: XbaI; This library
                     was supplied by Holly Cline (Cold Spring Harbor Labs).
                     cDNA synthesis with oligo dT Xba I (Xba I cloning site).
                     RNA: stage 50-56 tadpoles, total brain tissue, GTC
                     extraction method."
ORIGIN

Alignment Scores:
Pred. No.:                2.39e+03        Length:         417
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       1               Gaps:           0

US-09-858-332G-28 (1-6) x AW159517 (1-417)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        139 CATAATCACAACCACAAT 156

RESULT 291
BE606870
LOCUS       BE606870                417 bp    mRNA    linear   EST 22-AUG-2000
DEFINITION  WHE0902_A10_A20ZS Wheat 5-15 DAP spike cDNA library Triticum
            aestivum cDNA clone WHE0902_A10_A20, mRNA sequence.
ACCESSION   BE606870
VERSION     BE606870.1  GI:9884034
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 417)
  AUTHORS   Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D.,
            Han,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J.,
            Seaton,C.L. and Tong,J.C.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - 5-15 DAP spike cDNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene SK primer.
FEATURES             Location/Qualifiers
     source          1. .417
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /cultivar="Chinese Spring"
                     /db_xref="taxon:4565"
                     /clone="WHE0902_A10_A20"
                     /tissue_type="Spike"
                     /dev_stage="Adult plant"
                     /lab_host="E. coli SOLR"
                     /clone_lib="Wheat 5-15 DAP spike cDNA library"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; Plants were grown in the
                     greenhouse. Spikes at 5, 10 and 15 DAP were  harvested,
                     total RNA and poly(A) RNA were prepared, a cDNA library
                     was  made, and the cDNA clones were in vivo excised to
                     give pBluescript  phagemids in the TJ Close lab (Choi,
                     Close, Fenton) at the University  of California,
                     Riverside. Plasmid DNA preparations and DNA sequencing
                     were performed in the OD Anderson lab (all other
                     authors)."
ORIGIN

Alignment Scores:
Pred. No.:                2.39e+03        Length:         417
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       2               Gaps:           0

US-09-858-332G-28 (1-6) x BE606870 (1-417)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        278 CACAACCACAACCACAAT 295

RESULT 292
CN571761
LOCUS       CN571761                417 bp    mRNA    linear   EST 04-MAY-2004
DEFINITION  rg58h10.x1 Meloidogyne hapla female SMART pGEM Meloidogyne hapla
            cDNA 3' similar to contains MER30 repetitive element ;, mRNA
```

```
            sequence.
ACCESSION   CN571761
VERSION     CN571761.1  GI:46982449
KEYWORDS    EST.
SOURCE      Meloidogyne hapla
  ORGANISM  Meloidogyne hapla
            Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
            Tylenchoidea; Meloidogynidae; Meloidogyninae; Meloidogyne.
REFERENCE   1  (bases 1 to 417)
  AUTHORS   McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J.,
            Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B.,
            Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C.,
            Tsagareishvili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C.,
            Underwood,K., Steptoe,M., Allen,M., Person,B., Swaller,T.,
            Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     The Washington Univ. Nematode EST Project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: rg58h10.y1
            Contact: McCarter JP
            The Washington Univ. Nematode EST Project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Cloned unidirectionally. Poly(A)+ RNA was concentrated and purified
            using Dynabeads (Dynal) and mRNA eluted for first strand synthesis.
            First strand cDNA was created using MMLV RT (Powerscript, Clontech)
            and primed with oligo(dT) with XhoI site and 5'SMART 'anchor' added
            using chimeric DNA-RNA oligo. 12 PCR cycles were done using first
            strand and primers specific to SMART oligo and 3' end. Double
            stranded cDNA was digested using XhoI/NotI, fractioned on
            Chroma-spin 400 columns (Clontech) and ligated to digested
            pGEM-11zf(+) plasmid. Chemically competent DH10B cells were used as
            host cells. Library materials provided by Dr. David Bird of North
            Carolina State University. Library construction by Jeff Rousch. See
            www.nematode.net for additional project information.
            Seq primer: Primer name ambiguous.
FEATURES              Location/Qualifiers
     source           1. .417
                      /organism="Meloidogyne hapla"
                      /mol_type="mRNA"
                      /db_xref="taxon:6305"
                      /tissue_type="whole organism"
                      /dev_stage="adult"
                      /lab_host="DH10B"
                      /clone_lib="Meloidogyne hapla female SMART pGEM"
                      /note="Vector: plasmid (ampicillin resistant); Site_1:
                      XhoI; Site_2: NotI; Cloned unidirectionally. Poly(A)+ RNA
                      was concentrated and purified using Dynabeads (Dynal) and
                      mRNA eluted for first strand synthesis. First strand cDNA
                      was created using MMLV RT (Powerscript, Clontech) and
                      primed with oligo(dT) with XhoI site and 5'SMART 'anchor'
                      added using chimeric DNA-RNA oligo. 12 PCR cycles were
                      done using first strand and primers specific to SMART
                      oligo and 3' end. Double stranded cDNA was digested using
                      XhoI/NotI, fractioned on Chroma-spin 400 columns
                      (Clontech) and ligated to digested pGEM-11zf(+) plasmid.
                      Chemically competent DH10B cells were used as host cells.
                      Library materials provided by Dr. David Bird of North
                      Carolina State University. Library construction by Jeff
                      Rousch. See www.nematode.net for additional project
                      information."
ORIGIN

Alignment Scores:
Pred. No.:              2.39e+03        Length:         417
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     7               Gaps:           0
```

```
US-09-858-332G-28 (1-6) x CN571761 (1-417)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        142 CATAATCACAATCATAAT 159

RESULT 293
W98585
LOCUS       W98585                  417 bp    mRNA    linear   EST 16-JUL-1996
DEFINITION  mg09b04.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:423247 5' similar to gb:M38188 OVARIAN GRANULOSA CELL
            13.0 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION   W98585
VERSION     W98585.1  GI:1428575
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 417)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:257799
            Seq primer: ETPrimer
            High quality sequence stop: 342.
FEATURES              Location/Qualifiers
     source           1. .417
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:423247"
                      /sex="unknown"
                      /tissue_type="embryo"
                      /dev_stage="13.5-14.5dpc total fetus"
                      /lab_host="DH10B"
                      /clone_lib="Soares mouse embryo NbME13.5 14.5"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTTT
                      T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                      14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                      State Univ., from 2 ]; double-stranded cDNA was ligated to
                      Eco RI adaptors (Pharmacia), digested with Not I and
                      cloned into the Not I and Eco RI sites of the modified
                      pT7T3 vector. Library went through one round of
                      normalization, and was constructed by Bento Soares and
                      M.Fatima Bonaldo.  "
ORIGIN

Alignment Scores:
Pred. No.:              2.39e+03        Length:         417
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
```

```
DB:                    8              Gaps:          0

US-09-858-332G-28 (1-6) x W98585 (1-417)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        153 CACAACCATAACCACAAC 170

RESULT 294
CF096176
LOCUS       CF096176                419 bp    mRNA    linear   EST 22-JUL-2003
DEFINITION  QHN22F11.yg.ab1 QH_N sunflower H.argophyllus (drought stress)
            Helianthus argophyllus cDNA clone QHN22F11, mRNA sequence.
ACCESSION   CF096176
VERSION     CF096176.1  GI:33135243
KEYWORDS    EST.
SOURCE      Helianthus argophyllus
  ORGANISM  Helianthus argophyllus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; campanulids; Asterales; Asteraceae; Asteroideae;
            Heliantheae; Helianthus.
REFERENCE   1  (bases 1 to 419)
  AUTHORS   Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J.,
            Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y.,
            Lai,Z., Church,S., Jackson,L. and Bradford,K.
  TITLE     Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
            belongs to contig QH_CA_Contig6346, see http://cgpdb.ucdavis.edu/
            for details.
            Plate: QHN22  row: F  column: 11.
FEATURES              Location/Qualifiers
     source           1. .419
                      /organism="Helianthus argophyllus"
                      /mol_type="mRNA"
                      /db_xref="taxon:73275"
                      /clone="QHN22F11"
                      /lab_host="E.coli"
                      /clone_lib="QH_N sunflower H.argophyllus (drought stress)"
                      /note="Vector: pGEM-T; The library was constructed from
                      three different sources (seedling, root and leaf) of RNA
                      from a single genotype. cDNAs were pooled and cloned into
                      a high-copy vector pGEM-T. Details of library construction
                      can be obtained at http://cgpdb.ucdavis.edu/"
ORIGIN

Alignment Scores:
Pred. No.:              2.4e+03        Length:         419
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     6              Gaps:           0

US-09-858-332G-28 (1-6) x CF096176 (1-419)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        289 CATAACCATAATCATAAT 306

RESULT 295
CV546384
LOCUS       CV546384                419 bp    mRNA    linear   EST 22-OCT-2004
DEFINITION  SS2534.txt NaCl treated Chinese cabbage seedling library Brassica
            rapa subsp. pekinensis cDNA clone SS2534.txt 5', mRNA sequence.
ACCESSION   CV546384
VERSION     CV546384.1  GI:54420973
KEYWORDS    EST.
SOURCE      Brassica rapa subsp. pekinensis
  ORGANISM  Brassica rapa subsp. pekinensis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 419)
  AUTHORS   Kang,J.S., Ryu,S.H., Yang,K.A. and Lim,C.O.
  TITLE     Expressed Sequence Tags of NaCl treated Chinese cabbage seedlings
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Lim, C.O.
            Plant Molecular Biology & Biotechnology Research Centre
            Gyeongsang National University
            #900 Gazwa-dong, Jinju 660-701, Korea
            Tel: 82 55 751 6255
            Fax: 82 55 759 9363
            Email: colim@nongae.gsnu.ac.kr
            Seq primer: T7.
FEATURES              Location/Qualifiers
     source           1. .419
                      /organism="Brassica rapa subsp. pekinensis"
                      /mol_type="mRNA"
                      /cultivar="Jangwon"
                      /sub_species="pekinensis"
                      /db_xref="taxon:51351"
                      /clone="SS2534.txt"
                      /dev_stage="seedling"
                      /lab_host="XL-1 Blue"
                      /clone_lib="NaCl treated Chinese cabbage seedling library"
                      /note="Vector: pSPOT1; Site_1: Sal I; Site_2: Not I"
ORIGIN

Alignment Scores:
Pred. No.:              2.4e+03        Length:         419
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     7              Gaps:           0

US-09-858-332G-28 (1-6) x CV546384 (1-419)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        173 CACAACCACAACCACAAC 190

RESULT 296
AW985649
LOCUS       AW985649                420 bp    mRNA    linear   EST 02-JUN-2000
DEFINITION  uf84a09.y1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
            IMAGE:1528792 5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0
            KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION   AW985649
VERSION     AW985649.1  GI:8180089
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 420)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:944892
```

```
            Seq primer: -40RP from Gibco
            High quality sequence stop: 322.
FEATURES              Location/Qualifiers
     source           1. .420
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:1528792"
                      /sex="female (lactating)"
                      /tissue_type="mammary gland"
                      /lab_host="DH10B"
                      /clone_lib="Soares_mammary_gland_NMLMG"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; 1st strand cDNA was prepared from mammary
                      gland tissue from a lactating female, and was then primed
                      with a Not I - oligo(dT) primer. Double-stranded  cDNA was
                      ligated to Eco RI adaptors (Pharmacia), digested with Not
                      I  and cloned into the Not I and Eco RI sites of the
                      modified pT7T3 vector.  Library is normalized.  Library
                      was constructed by Bento Soares and M. Fatima Bonaldo. "
ORIGIN

Alignment Scores:
Pred. No.:              2.4e+03         Length:         420
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AW985649 (1-420)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        212 CACAACCATAACCACAAC 229

RESULT 297
CF991917
LOCUS       CF991917                420 bp    mRNA    linear   EST 04-FEB-2005
DEFINITION  23707rsicee_4196.y1 Oryza sativa cv. LYP9 tillering whole plant
            cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF991917
VERSION     CF991917.1  GI:58593609
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 420)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 420
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .420
                      /organism="Oryza sativa (indica cultivar-group)"
                      /mol_type="mRNA"
                      /cultivar="LYP9"
                      /db_xref="taxon:39946"
                      /tissue_type="whole plant"
                      /dev_stage="tillering"
                      /clone_lib="Oryza sativa cv. LYP9 tillering whole plant
                      cDNA library"
ORIGIN

Alignment Scores:
Pred. No.:              2.4e+03         Length:         420
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     7               Gaps:           0

US-09-858-332G-28 (1-6) x CF991917 (1-420)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        185 CACAACCACAACCACAAC 202

RESULT 298
Z34015
LOCUS       Z34015                  420 bp    mRNA    linear   EST 06-JUN-1994
DEFINITION  ATTS3063 Perp-dry-seedA Arabidopsis thaliana cDNA clone PAP434 5',
            mRNA sequence.
ACCESSION   Z34015
VERSION     Z34015.1  GI:496772
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 420)
  AUTHORS   CNRS.
  TITLE     The Arabidopsis thaliana transcribed genome: the GDR cDNA program
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Raynal M., Grellet F., Laudie M., Meyer Y., Cooke R.,
            Delseny M.
            Laboratoire de Physiologie et Biologie Moleculaires Vegetales, URA
            CNRS 565
            Universite de Perpignan
            Avenue de Villeneuve,66860 Perpignan Cedex,France
            Email: raynal@univ-perp.fr.
FEATURES              Location/Qualifiers
     source           1. .420
                      /organism="Arabidopsis thaliana"
                      /mol_type="mRNA"
                      /ecotype="Columbia"
                      /db_xref="taxon:3702"
                      /clone="PAP434"
                      /clone_lib="Perp-dry-seedA"
                      /note="Vector: Lambda ZAPII oriented; Site_1: EcoRI;
                      Site_2: XhoI; tissue_type: Dry seeds "
ORIGIN

Alignment Scores:
Pred. No.:              2.4e+03         Length:         420
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     8               Gaps:           0

US-09-858-332G-28 (1-6) x Z34015 (1-420)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         99 CACAACCACAACCACAAT 116
```

```
RESULT 299
AQ655474/c
LOCUS       AQ655474               420 bp    DNA     linear   GSS 22-JUN-1999
DEFINITION  Sheared DNA-25D5.TF Sheared DNA Trypanosoma brucei genomic clone
            Sheared DNA-25D5, genomic survey sequence.
ACCESSION   AQ655474
VERSION     AQ655474.1  GI:5148660
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 420)
  AUTHORS   El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
            Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S.,
            Donelson,J., Fraser,C. and Adams,M.
  TITLE     Determination of clone end sequences from Trypanosoma brucei GUTat
            10.1 sheared DNA library
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Najib M. El-Sayed
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: nelsayed@tigr.org
            Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
            DNA library constructed at TIGR. Clones will be available for
            distribution through ATCC. Sheared DNA end sequences search page:
            http://www.tigr.org/tdb/mdb/tbdb/.
            Seq primer: M13-Forward
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .420
                     /organism="Trypanosoma brucei"
                     /mol_type="genomic DNA"
                     /strain="TREU927/4 GUTat 10.1"
                     /db_xref="taxon:5691"
                     /clone="Sheared DNA-25D5"
                     /clone_lib="Sheared DNA"
                     /note="Vector: pUC18; Site_1: SmaI; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from a cloned population of
                     Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
                     sheared to give a tight size distribution (approx 2 kb).
                     The v + i method used for the library construction is
                     described in detail in Smith, H.O. and Venter, J.C.
                     (Making small insert libraries for whole genome shotgun
                     sequencing projects.  In Genome Sequencing: A Practical
                     Approach, eds. M. Vaudin and B. Barell, Oxford University
                     Press, 1999)."
ORIGIN

Alignment Scores:
Pred. No.:                2.4e+03        Length:          420
Score:                    42.00          Matches:         6
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       9              Gaps:            0

US-09-858-332G-28 (1-6) x AQ655474 (1-420)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        341 CACAATCACAACCACAAC 324

RESULT 300
AV406807
LOCUS       AV406807               421 bp    mRNA    linear   EST 10-MAR-2004
DEFINITION  AV406807 Lotus japonicus young plants (two-week old) Lotus
            corniculatus var. japonicus cDNA clone MWL010a05_r 5', mRNA
            sequence.
ACCESSION   AV406807
VERSION     AV406807.1  GI:7719661
KEYWORDS    EST.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1  (bases 1 to 421)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Generation of 7137 non-redundant expressed sequence tags from a
            legume, Lotus japonicus
  JOURNAL   DNA Res. 7 (2), 127-130 (2000)
  PUBMED    10819328
COMMENT     Contact: Yasukazu Nakamura
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: ynakamu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .421
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="mRNA"
                     /isolate="Miyakojima MG-20"
                     /db_xref="taxon:34305"
                     /clone="MWL010a05_r"
                     /tissue_type="whole plant"
                     /dev_stage="two-week old plant"
                     /clone_lib="Lotus japonicus young plants (two-week old)"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI; synonym: Lotus japonicus"
ORIGIN

Alignment Scores:
Pred. No.:                2.41e+03       Length:          421
Score:                    42.00          Matches:         6
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       1              Gaps:            0

US-09-858-332G-28 (1-6) x AV406807 (1-421)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        146 CACAATCACAATCACAAT 163

RESULT 301
BF595803/c
LOCUS       BF595803               421 bp    mRNA    linear   EST 13-JUL-2004
DEFINITION  sv04g09.y1 Gm-c1056 Glycine soja cDNA clone GENOME SYSTEMS CLONE
            ID: Gm-c1056-1074 5' similar to TR:Q9SY30 Q9SY30 T17H7.16. ;, mRNA
            sequence.
ACCESSION   BF595803
VERSION     BF595803.1  GI:11688127
KEYWORDS    EST.
SOURCE      Glycine soja
  ORGANISM  Glycine soja
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 421)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,
            Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
            Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
            Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
            Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
```

```
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            When it has been determined, an EST from the other end of this
            clone is listed in the 'Other ESTs on clone' field. This clone is
            available through: Biogenetic Services, 801 32nd Ave. Brookings, SD
            57006 USA (phone: 800 423 4163; email: info@biogeneticservices.com)
            Insert Length: 893   Std Error: 0.00
            High quality sequence stop: 421.
FEATURES              Location/Qualifiers
     source           1. .421
                      /organism="Glycine soja"
                      /mol_type="mRNA"
                      /db_xref="taxon:3848"
                      /clone="GENOME SYSTEMS CLONE ID: Gm-c1056-1074"
                      /tissue_type="Whole seedling, 4 day old"
                      /lab_host="DH10B"
                      /clone_lib="Gm-c1056"
                      /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                      XhoI; The cDNA library was constructed from mRNA isolated
                      from 4 day old seedling of PI468916. The seedlings were
                      germinated in a growth chamber using germination paper.
                      Complementary DNA was synthesized from mRNA using a primer
                      consisting of a poly(dT) sequence with a XhoI restriction
                      site. EcoRI adapters were ligated to the blunt-ended cDNA
                      fragments followed by XhoI digestion. The cDNA fragments
                      were directionally cloned into the EcoRI-XhoI restriction
                      site of the pBluescript vector. The ligated cDNA fragments
                      were transformed into DH10B host cells (GibcoBRL). This
                      library was constructed in the  laboratory of Dr. Randy
                      Shoemaker."
ORIGIN

Alignment Scores:
Pred. No.:              2.41e+03        Length:         421
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-858-332G-28 (1-6) x BF595803 (1-421)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         43 CACAATCACAATCACAAT 26

RESULT 302
BU809035
LOCUS       BU809035                421 bp    mRNA    linear   EST 15-OCT-2002
DEFINITION  UL52PH05 Populus leaf cDNA library Populus tremula x Populus
            tremuloides cDNA 5 prime, mRNA sequence.
ACCESSION   BU809035
VERSION     BU809035.1  GI:23959477
KEYWORDS    EST.
SOURCE      Populus tremula x Populus tremuloides
  ORGANISM  Populus tremula x Populus tremuloides
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Malpighiales; Salicaceae; Saliceae; Populus.
REFERENCE   1  (bases 1 to 421)
  AUTHORS   Unneberg,P., Bhalerao,R.R., Jansson,S. and Sterky,F.
  TITLE     The poplar tree transcriptome: Analysis of expressed sequence tags
            from multiple libraries
  JOURNAL   Unpublished (2002)
COMMENT     Contact: BHALERAO RUPALI R.
            Umea Plant Science Center
            Department of Plant Physiology
            University of Umea, 901 87 Umea, Sweden
            Tel: +46 90 786 5279
            Fax: +46 90 786 6676
            Email: rupali.bhalerao@plantphys.umu.se.
FEATURES              Location/Qualifiers
     source           1. .421
                      /organism="Populus tremula x Populus tremuloides"
                      /mol_type="mRNA"
                      /db_xref="taxon:47664"
                      /clone_lib="Populus leaf cDNA library"
                      /note="Organ: leaf"
ORIGIN

Alignment Scores:
Pred. No.:              2.41e+03        Length:         421
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     5               Gaps:           0

US-09-858-332G-28 (1-6) x BU809035 (1-421)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        160 CACAACCACAACCACAAC 177

RESULT 303
AV768209
LOCUS       AV768209                422 bp    mRNA    linear   EST 18-AUG-2004
DEFINITION  AV768209 Lotus japonicus Young plants (two-weeks old) Lotus
            corniculatus var. japonicus cDNA clone MWM228b05_f 3', mRNA
            sequence.
ACCESSION   AV768209
VERSION     AV768209.1  GI:45351482
KEYWORDS    EST.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1  (bases 1 to 422)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Characteristics of the Lotus japonicus gene repertoire deduced from
            large-scale expressed sequence tag (EST) analysis
  JOURNAL   Plant Mol. Biol. 54 (3), 405-414 (2004)
   PUBMED   15284495
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES              Location/Qualifiers
     source           1. .422
                      /organism="Lotus corniculatus var. japonicus"
                      /mol_type="mRNA"
                      /variety="japonicus"
                      /db_xref="taxon:34305"
                      /clone="MWM228b05_f"
                      /tissue_type="Young plants (two-weeks old)"
                      /clone_lib="Lotus japonicus Young plants (two-weeks old)"
                      /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                      XhoI; isolate=Miyakojima MG-20"
ORIGIN

Alignment Scores:
Pred. No.:              2.42e+03        Length:         422
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0
```

US-09-858-332G-28 (1-6) x AV768209 (1-422)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        355 CATAATCATAATCATAAT 372
```

RESULT 304
BG262466
LOCUS BG262466 422 bp mRNA linear EST 16-FEB-2001
DEFINITION WHE0936_C12_F24ZS Wheat 5-15 DAP spike cDNA library Triticum aestivum cDNA clone WHE0936_C12_F24, mRNA sequence.
ACCESSION BG262466
VERSION BG262466.1 GI:12863654
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 422)
AUTHORS Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L. and Tong,J.C.
TITLE The structure and function of the expressed portion of the wheat genomes - 5-15 DAP spike cDNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oandersn@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
Seq primer: Stratagene SK primer.

FEATURES Location/Qualifiers
source 1. .422
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE0936_C12_F24"
/tissue_type="Spike"
/dev_stage="Adult plant"
/lab_host="E. coli SOLR"
/clone_lib="Wheat 5-15 DAP spike cDNA library"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site_1: EcoRI; Site_2: XhoI; Plants were grown in the greenhouse. Spikes at 5, 10 and 15 DAP were harvested, total RNA and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give pBluescript phagemids in the TJ Close lab (Choi, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

ORIGIN

Alignment Scores:

Pred. No.:	2.42e+03	Length:	422
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-28 (1-6) x BG262466 (1-422)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        309 CACAACCACAACCACAAT 326
```

RESULT 305
CF979659/c
LOCUS CF979659 422 bp mRNA linear EST 25-NOV-2003
DEFINITION rg79e02.y1 Meloidogyne hapla female SL1 pGEM Meloidogyne hapla cDNA 5', mRNA sequence.
ACCESSION CF979659
VERSION CF979659.1 GI:38512708
KEYWORDS EST.
SOURCE Meloidogyne hapla
ORGANISM Meloidogyne hapla
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina; Tylenchoidea; Meloidogynidae; Meloidogyninae; Meloidogyne.
REFERENCE 1 (bases 1 to 422)
AUTHORS McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagareishvili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe,M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE The Washington Univ. Nematode EST Project, 1999
JOURNAL Unpublished (1999)
COMMENT Contact: McCarter JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Cloned unidirectionally. Poly (A)+ RNA was concentrated and purified using Dynabeads (Dynal) and mRNA eluted for 1ST strand synthesis. 1st strand cDNA created using MMLV RT(PowerScript, CLontech) and primed with oligo(dT) with XhoI site (primer CDSIII/3-XhoI) and 5' SMART anchor added using chimeric DNA-RNA oligo (SMART-NotI-r-GGG). SMART-NotI-r-GGG was used in case a PCR using the SMART template was also desired. 15 PCR cycles were done using 1st strand and primers specific to SL1 leader sequence (SL1-NotI) and 3' end (XhoI-No-dT). dscDNA was digested using XhoI/NotI, fractionated on Chroma-spin 400 columns (Clontech) and ligated to digested pGEM -11zf(+) plasmid. Chemically competent cells were used as host. This library was provided by David Bird's lab (david_bird@ncsu.edu), of North Carolina State University.
Putative full length read
The vector to vector length is 444
Seq primer: Sp6.

FEATURES Location/Qualifiers
source 1. .422
/organism="Meloidogyne hapla"
/mol_type="mRNA"
/db_xref="taxon:6305"
/sex="female"
/tissue_type="whole organism"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="Meloidogyne hapla female SL1 pGEM"
/note="Vector: pGEM-11zf(+) (Promega); Site_1: XhoI; Site_2: NotI; Cloned unidirectionally. Poly (A)+ RNA was concentrated and purified using Dynabeads (Dynal) and mRNA eluted for 1ST strand synthesis. 1st strand cDNA created using MMLV RT(PowerScript, CLontech) and primed with oligo(dT) with XhoI site (primer CDSIII/3-XhoI) and 5' SMART anchor added using chimeric DNA-RNA oligo (SMART-NotI-r-GGG). SMART-NotI-r-GGG was used in case a PCR using the SMART template was also desired. 15 PCR cycles were done using 1st strand and primers specific to SL1 leader sequence (SL1-NotI) and 3' end (XhoI-No-dT). dscDNA was digested using XhoI/NotI, fractionated on Chroma-spin 400 columns (Clontech) and ligated to digested pGEM -11zf(+) plasmid. Chemically competent cells were used as host. This library was provided by David Bird's lab (david_bird@ncsu.edu), of North Carolina State University."

ORIGIN

```
Alignment Scores:
Pred. No.:                2.42e+03        Length:         422
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       7               Gaps:           0

US-09-858-332G-28 (1-6) x CF979659 (1-422)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        255 CATAATCACAATCATAAT 238

RESULT 306
AV420924
LOCUS       AV420924                423 bp    mRNA    linear   EST 10-MAR-2004
DEFINITION  AV420924 Lotus japonicus young plants (two-week old) Lotus
            corniculatus var. japonicus cDNA clone MWM188g08_r 5', mRNA
            sequence.
ACCESSION   AV420924
VERSION     AV420924.1  GI:7774023
KEYWORDS    EST.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1  (bases 1 to 423)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Generation of 7137 non-redundant expressed sequence tags from a
            legume, Lotus japonicus
  JOURNAL   DNA Res. 7 (2), 127-130 (2000)
   PUBMED   10819328
COMMENT     Contact: Yasukazu Nakamura
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: ynakamu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .423
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="mRNA"
                     /isolate="Miyakojima MG-20"
                     /db_xref="taxon:34305"
                     /clone="MWM188g08_r"
                     /tissue_type="whole plant"
                     /dev_stage="two-week old plant"
                     /clone_lib="Lotus japonicus young plants (two-week old)"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI; synonym: Lotus japonicus"
ORIGIN

Alignment Scores:
Pred. No.:                2.42e+03        Length:         423
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       1               Gaps:           0

US-09-858-332G-28 (1-6) x AV420924 (1-423)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        107 CACAACCACAACCACAAC 124

RESULT 307
CN194535/c
LOCUS       CN194535                423 bp    mRNA    linear   EST 05-APR-2004
DEFINITION  rg95g07.y1 Meloidogyne hapla female SMART pGEM Meloidogyne hapla
            cDNA 5', mRNA sequence.
ACCESSION   CN194535
VERSION     CN194535.1  GI:46219474
KEYWORDS    EST.
SOURCE      Meloidogyne hapla
  ORGANISM  Meloidogyne hapla
            Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
            Tylenchoidea; Meloidogynidae; Meloidogyninae; Meloidogyne.
REFERENCE   1  (bases 1 to 423)
  AUTHORS   McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J.,
            Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B.,
            Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C.,
            Tsagareishvili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C.,
            Underwood,K., Steptoe,M., Allen,M., Person,B., Swaller,T.,
            Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     The Washington Univ. Nematode EST Project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: McCarter JP
            The Washington Univ. Nematode EST Project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Cloned unidirectionally. Poly(A)+ RNA was concentrated and purified
            using Dynabeads (Dynal) and mRNA eluted for first strand synthesis.
            First strand cDNA was created using MMLV RT (Powerscript, Clontech)
            and primed with oligo(dT) with XhoI site and 5'SMART 'anchor' added
            using chimeric DNA-RNA oligo. 12 PCR cycles were done using first
            strand and primers specific to SMART oligo and 3' end. Double
            stranded cDNA was digested using XhoI/NotI, fractioned on
            Chroma-spin 400 columns (Clontech) and ligated to digested
            pGEM-11zf(+) plasmid. Chemically competent DH10B cells were used as
            host cells. Library materials provided by Dr. David Bird of North
            Carolina State University. Library construction by Jeff Rousch. See
            www.nematode.net for additional project information.
            Seq primer: Sp6
            High quality sequence stop: 423.
FEATURES             Location/Qualifiers
     source          1. .423
                     /organism="Meloidogyne hapla"
                     /mol_type="mRNA"
                     /db_xref="taxon:6305"
                     /tissue_type="whole organism"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="Meloidogyne hapla female SMART pGEM"
                     /note="Vector: plasmid (ampicillin resistant); Site_1:
                     XhoI; Site_2: NotI; Cloned unidirectionally. Poly(A)+ RNA
                     was concentrated and purified using Dynabeads (Dynal) and
                     mRNA eluted for first strand synthesis. First strand cDNA
                     was created using MMLV RT (Powerscript, Clontech) and
                     primed with oligo(dT) with XhoI site and 5'SMART 'anchor'
                     added using chimeric DNA-RNA oligo. 12 PCR cycles were
                     done using first strand and primers specific to SMART
                     oligo and 3' end. Double stranded cDNA was digested using
                     XhoI/NotI, fractioned on Chroma-spin 400 columns
                     (Clontech) and ligated to digested pGEM-11zf(+) plasmid.
                     Chemically competent DH10B cells were used as host cells.
                     Library materials provided by Dr. David Bird of North
                     Carolina State University. Library construction by Jeff
                     Rousch. See www.nematode.net for additional project
                     information."
ORIGIN

Alignment Scores:
Pred. No.:                2.42e+03        Length:         423
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       7               Gaps:           0
```

US-09-858-332G-28 (1-6) x CN194535 (1-423)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        259 CATAATCACAATCATAAT 242
```

RESULT 308
AV809042/c
LOCUS AV809042 424 bp mRNA linear EST 29-MAR-2002
DEFINITION AV809042 RAFL9 Arabidopsis thaliana cDNA clone RAFL09-57-N16 3', mRNA sequence.
ACCESSION AV809042
VERSION AV809042.1 GI:19843027
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 424)
AUTHORS Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M., Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y., Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y. and Shinozaki,K.
TITLE Large scale analysis of Arabidopsis full-length cDNA (2002b)
JOURNAL Unpublished (2002)
COMMENT Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@rtc.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998). cDNA cleaved with BamHI and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et al., submitted for publication) digested with BamHI and SalI. This clone is in a modified pBluescript vector. Please visit our web site (http://www.gsc.riken.go.jp/e/plant/index _e.html) for further details.

FEATURES Location/Qualifiers
source 1. .424
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAFL09-57-N16"
/dev_stage="plants at various developmental stages from germination to mature seeds"
/lab_host="DH10B"
/clone_lib="RAFL9"
/note="Site_1: BamHI; Site_2: SalI; subjected to dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24 hr) treatments"

ORIGIN

Alignment Scores:

Pred. No.:	2.43e+03	Length:	424
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0

US-09-858-332G-28 (1-6) x AV809042 (1-424)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        308 CACAACCACAACCACAAC 291
```

RESULT 309
CF369371/c
LOCUS CF369371 424 bp mRNA linear EST 27-AUG-2003
DEFINITION rg58h10.y1 Meloidogyne hapla female SMART pGEM Meloidogyne hapla cDNA 5', mRNA sequence.
ACCESSION CF369371
VERSION CF369371.1 GI:34316617
KEYWORDS EST.
SOURCE Meloidogyne hapla
ORGANISM Meloidogyne hapla
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina; Tylenchoidea; Meloidogynidae; Meloidogyninae; Meloidogyne.
REFERENCE 1 (bases 1 to 424)
AUTHORS McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagareishvili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe,M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE The Washington Univ. Nematode EST Project, 1999
JOURNAL Unpublished (1999)
COMMENT Contact: McCarter JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Cloned unidirectionally. Poly(A)+ RNA was concentrated and purified using Dynabeads (Dynal) and mRNA eluted for first strand synthesis. First strand cDNA was created using MMLV RT (Powerscript, Clontech) and primed with oligo(dT) with XhoI site and 5'SMART 'anchor' added using chimeric DNA-RNA oligo. 12 PCR cycles were done using first strand and primers specific to SMART oligo and 3' end. Double stranded cDNA was digested using XhoI/NotI, fractioned on Chroma-spin 400 columns (Clontech) and ligated to digested pGEM-11zf(+) plasmid. Chemically competent DH10B cells were used as host cells. Library materials provided by Dr. David Bird of North Carolina State University. Library construction by Jeff Rousch. See www.nematode.net for additional project information.
Putative full length read
The vector to vector length is 439
Seq primer: Sp6.

FEATURES Location/Qualifiers
source 1. .424
/organism="Meloidogyne hapla"
/mol_type="mRNA"
/db_xref="taxon:6305"
/tissue_type="whole organism"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="Meloidogyne hapla female SMART pGEM"
/note="Vector: plasmid (ampicillin resistant); Site_1: XhoI; Site_2: NotI; Cloned unidirectionally. Poly(A)+ RNA was concentrated and purified using Dynabeads (Dynal) and mRNA eluted for first strand synthesis. First strand cDNA was created using MMLV RT (Powerscript, Clontech) and primed with oligo(dT) with XhoI site and 5'SMART 'anchor' added using chimeric DNA-RNA oligo. 12 PCR cycles were done using first strand and primers specific to SMART oligo and 3' end. Double stranded cDNA was digested using XhoI/NotI, fractioned on Chroma-spin 400 columns (Clontech) and ligated to digested pGEM-11zf(+) plasmid. Chemically competent DH10B cells were used as host cells. Library materials provided by Dr. David Bird of North Carolina State University. Library construction by Jeff Rousch. See www.nematode.net for additional project information."

ORIGIN

Alignment Scores:

Pred. No.:	2.43e+03	Length:	424
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0

```
Query Match:            100.00%       Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-28 (1-6) x CF369371 (1-424)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        261 CATAATCACAATCATAAT 244

RESULT 310
AQ171001
LOCUS       AQ171001                424 bp    DNA     linear   GSS 17-OCT-1998
DEFINITION  HS_2217_A2_D10_MF CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2217 Col=20 Row=G, genomic survey
            sequence.
ACCESSION   AQ171001
VERSION     AQ171001.1  GI:3568368
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 424)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 2217  row: G  column: 20
            Class: BAC ends
            High quality sequence stop: 424.
FEATURES             Location/Qualifiers
     source          1. .424
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=2217 Col=20 Row=G"
                     /sex="male"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
ORIGIN

Alignment Scores:
Pred. No.:              2.43e+03      Length:        424
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     9             Gaps:          0

US-09-858-332G-28 (1-6) x AQ171001 (1-424)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        106 CACAACCATAATCACAAT 123

RESULT 311
BQ576463
LOCUS       BQ576463                425 bp    mRNA    linear   EST 05-JUL-2004
DEFINITION  sap17e08.y1 Gm-c1082 Glycine max cDNA clone SOYBEAN CLONE ID:
            Gm-c1082-2896 5' similar to TR:O65168 O65168 AUXIN-INDUCED PROTEIN
            ;, mRNA sequence.
ACCESSION   BQ576463
VERSION     BQ576463.1  GI:21479780
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 425)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,
            Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
            Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
            Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
            Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            When it has been determined, an EST from the other end of this
            clone is listed in the 'Other ESTs on clone' field. This clone is
            available through: Biogenetic Services, 801 32nd Ave. Brookings, SD
            57006 USA (phone: 800 423 4163; email: info@biogeneticservices.com)
            Putative full length read vector to vector length is 501
            Seq primer: -40RP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .425
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /cultivar="Minsoy x Noir RI progeny"
                     /db_xref="taxon:3847"
                     /clone="SOYBEAN CLONE ID: Gm-c1082-2896"
                     /tissue_type="Roots of bulked Minsoy x Noir RI progeny"
                     /lab_host="DH10B"
                     /clone_lib="Gm-c1082"
                     /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                     XhoI; The mRNA was isolated from roots of bulked Minsoy x
                     Noir RI progeny that are phenotypically susceptible to
                     flooding damage. The plants were flooded (submerged) for
                     24 hours prior to harvest. Dr. Tara VanToai generously
                     provided the isolated mRNA. The roots were flash- frozen
                     in liquid nitrogen. Stratagene's cDNA Synthesis Kit
                     (catalog number 200401) was used to synthesize the cDNA.
                     First-strand synthesis was performed with 5-methyl dCTP,
                     hence the ligated cDNA was hemimethylated. A modification
                     of Stratagene's first-strand synthesis primer was used. An
                     'anchor' nucleotide (V=A, C, or G) was added to the 3' end
                     of the primer [GAGAGAGAGAGAGAGAGAACTAGTCTCGAG(T)18V] to
                     anchor the primer at the 5' end of the poly(A) tract.
                     After second-strand synthesis, the cDNA ends were filled
                     in with cloned Pfu DNA, ligated to EcoRI adapters and
                     subsequently phosphorylated. The cDNA was then
                     precipitated and redissolved in sterile, RNase-,
                     DNase-free water. The XhoI site within the first- strand
                     synthesis primer was then restricted by digestion with
                     XhoI from Promega (40U/ul); all XhoI sites in the cDNA
                     would be protected by their hemimethylated status. The
                     cDNA constructs were size-fractionated with a 500bp
                     cutoff, using Sephacryl S-500 High Resolution (Pharmacia
                     Biotech) in a 2-mm diameter column and a bed volume of
                     approximately 1ml. The column eluent was precipitated,
                     redissolved, and ligated into Stratagene's pBluescript II
                     XR Predigested vector (pBluescript II SK(+) vector that
                     has been digested with EcoRI and XhoI, and phosphorylated
                     by Stratagene). This library was constucted in the
                     laboratory of Dr. Paul Keim and Dr. Virginia H. Coryell at
```

Northern Arizona University."
ORIGIN

Alignment Scores:

Pred. No.:	2.43e+03	Length:	425
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	5	Gaps:	0

US-09-858-332G-28 (1-6) x BQ576463 (1-425)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        255 CATAACCATAACCATAAC 272
```

RESULT 312
CN193561/c
LOCUS CN193561 425 bp mRNA linear EST 05-APR-2004
DEFINITION rg88g02.y1 Meloidogyne hapla female SMART pGEM Meloidogyne hapla cDNA 5', mRNA sequence.
ACCESSION CN193561
VERSION CN193561.1 GI:46218500
KEYWORDS EST.
SOURCE Meloidogyne hapla
ORGANISM Meloidogyne hapla
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina; Tylenchoidea; Meloidogynidae; Meloidogyninae; Meloidogyne.
REFERENCE 1 (bases 1 to 425)
AUTHORS McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagareishvili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe,M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE The Washington Univ. Nematode EST Project, 1999
JOURNAL Unpublished (1999)
COMMENT Contact: McCarter JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Cloned unidirectionally. Poly(A)+ RNA was concentrated and purified using Dynabeads (Dynal) and mRNA eluted for first strand synthesis. First strand cDNA was created using MMLV RT (Powerscript, Clontech) and primed with oligo(dT) with XhoI site and 5'SMART 'anchor' added using chimeric DNA-RNA oligo. 12 PCR cycles were done using first strand and primers specific to SMART oligo and 3' end. Double stranded cDNA was digested using XhoI/NotI, fractioned on Chroma-spin 400 columns (Clontech) and ligated to digested pGEM-11zf(+) plasmid. Chemically competent DH10B cells were used as host cells. Library materials provided by Dr. David Bird of North Carolina State University. Library construction by Jeff Rousch. See www.nematode.net for additional project information.
Putative full length read
The vector to vector length is 440
Seq primer: Sp6.
FEATURES Location/Qualifiers
source 1. .425
/organism="Meloidogyne hapla"
/mol_type="mRNA"
/db_xref="taxon:6305"
/tissue_type="whole organism"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="Meloidogyne hapla female SMART pGEM"
/note="Vector: plasmid (ampicillin resistant); Site_1: XhoI; Site_2: NotI; Cloned unidirectionally. Poly(A)+ RNA was concentrated and purified using Dynabeads (Dynal) and mRNA eluted for first strand synthesis. First strand cDNA was created using MMLV RT (Powerscript, Clontech) and primed with oligo(dT) with XhoI site and 5'SMART 'anchor' added using chimeric DNA-RNA oligo. 12 PCR cycles were done using first strand and primers specific to SMART oligo and 3' end. Double stranded cDNA was digested using XhoI/NotI, fractioned on Chroma-spin 400 columns (Clontech) and ligated to digested pGEM-11zf(+) plasmid. Chemically competent DH10B cells were used as host cells. Library materials provided by Dr. David Bird of North Carolina State University. Library construction by Jeff Rousch. See www.nematode.net for additional project information."
ORIGIN

Alignment Scores:

Pred. No.:	2.43e+03	Length:	425
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	7	Gaps:	0

US-09-858-332G-28 (1-6) x CN193561 (1-425)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        261 CATAATCACAATCATAAT 244
```

RESULT 313
AI008498/c
LOCUS AI008498 426 bp mRNA linear EST 25-JAN-1999
DEFINITION EST202949 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone REMAY09 3' end, mRNA sequence.
ACCESSION AI008498
VERSION AI008498.1 GI:3222330
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 426)
AUTHORS Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J., Kerlavage,A.R. and Adams,M.D.
TITLE Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat Gene Index
JOURNAL Unpublished (1998)
COMMENT Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
FEATURES Location/Qualifiers
source 1. .426
/organism="Rattus sp."
/mol_type="mRNA"
/db_xref="ATCC (inhost):2016683"
/db_xref="taxon:10118"
/clone="REMAY09"
/dev_stage="embryo 8, 12, 18 dpc"
/clone_lib="Normalized rat embryo, Bento Soares"
/note="Vector: pT7T3Pac; Site_1: EcoRI; Site_2: NotI"
ORIGIN

Alignment Scores:

Pred. No.:	2.44e+03	Length:	426
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0

```
DB:                    1              Gaps:          0

US-09-858-332G-28 (1-6) x AI008498 (1-426)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        322 CACAACCATAACCACAAC 305

RESULT 314
BI653112
LOCUS       BI653112                426 bp    mRNA    linear   EST 12-SEP-2001
DEFINITION  603300651F1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5341069 5',
            mRNA sequence.
ACCESSION   BI653112
VERSION     BI653112.1  GI:15567348
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 426)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11866  row: f  column: 14
            High quality sequence stop: 424.
FEATURES             Location/Qualifiers
     source          1. .426
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="129,C57BL/6J,FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5341069"
                     /tissue_type="tumor, gross tissue"
                     /dev_stage="10 months"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Mam3"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigators
                     providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
                     Reference for transgenic model: Xu et al., Nature Genetics
                     22, 37-43 (1999)."
ORIGIN

Alignment Scores:
Pred. No.:              2.44e+03       Length:         426
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     3              Gaps:           0

US-09-858-332G-28 (1-6) x BI653112 (1-426)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        353 CACAACCATAACCACAAC 370

RESULT 315
BI798061
LOCUS       BI798061                426 bp    mRNA    linear   EST 02-OCT-2001
DEFINITION  H096C06 Endosperm library from Oryza sativa (10 days after
            anthesis) Oryza sativa cDNA clone H096C06, mRNA sequence.
ACCESSION   BI798061
VERSION     BI798061.1  GI:15849785
KEYWORDS    EST.
SOURCE      Oryza sativa
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 426)
  AUTHORS   Dong,H.T., Li,D.B., Zhuang,X.F., Dai,C.G., Sun,L.X., Pei,Y.X.,
            Wu,H.F., Jiang,Y.X., Yu,F.C., Gao,Q.K. and Lou,Y.C.
  TITLE     A Gene Expression Screen in Oryza sativa
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Haitao Dong, Debao Li
            Bioinfomatics and Gene Network Research Group
            Zhejiang University
            Kaixuan Road 268#, Hangzhou, Zhejiang, P.R.China
            Tel: 0086-571-86892051
            Fax: 0086-571-86961525
            Email: webmaster@estarray.org, URL: http://www.estarray.org
            Seq primer: M13 forward primer.
FEATURES             Location/Qualifiers
     source          1. .426
                     /organism="Oryza sativa"
                     /mol_type="mRNA"
                     /db_xref="taxon:4530"
                     /clone="H096C06"
                     /tissue_type="Endosperm"
                     /dev_stage="10 days after anthesis"
                     /clone_lib="Endosperm library from Oryza sativa (10 days
                     after anthesis)"
                     /note="Vector: pSport2"
ORIGIN

Alignment Scores:
Pred. No.:              2.44e+03       Length:         426
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     3              Gaps:           0

US-09-858-332G-28 (1-6) x BI798061 (1-426)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         49 CACAACCACAACCACAAC 66

RESULT 316
CO328401
LOCUS       CO328401                426 bp    mRNA    linear   EST 28-JUN-2004
DEFINITION  EP09137.3prime Exelixis FlyTag CK02 pCDNA-SK+ Drosophila
            melanogaster cDNA clone EP09137 3, mRNA sequence.
ACCESSION   CO328401
VERSION     CO328401.1  GI:49386835
KEYWORDS    EST.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 426)
  AUTHORS   Platt,D., Kopczynski,C., Muzong,C., Laufer,A., Leung,W.,
            Peterson,E. and Swimmer,C.
  TITLE     Exelixis FlyTag EST Project CK02 Library
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
```

```
           Plate: EP.91  row: D  column: 1
           High quality sequence stop: 388.
FEATURES             Location/Qualifiers
     source          1. .426
                     /organism="Drosophila melanogaster"
                     /mol_type="mRNA"
                     /db_xref="taxon:7227"
                     /clone="EP09137"
                     /clone_lib="Exelixis FlyTag CK02 pCDNA-SK+"
                     /note="Organ: mixed stage embryos, imaginal disks, and
                     adult heads; Vector: pCDNA-SK+; Site_1: NotI; Site_2:
                     XhoI; Random primed, normalized library from mixed stage
                     embryos, imaginal disks, and adult heads. Subset of
                     Exelixis FlyTag CK01 clones sequenced from 3' end"
ORIGIN

Alignment Scores:
Pred. No.:                2.44e+03       Length:         426
Score:                    42.00          Matches:        6
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       7              Gaps:           0

US-09-858-332G-28 (1-6) x CO328401 (1-426)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        140 CATAATCATAATCATAAT 157

RESULT 317
AU033706
LOCUS       AU033706                  427 bp    mRNA    linear   EST 28-APR-1999
DEFINITION  AU033706 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
            discoideum cDNA clone SLB340, mRNA sequence.
ACCESSION   AU033706
VERSION     AU033706.1  GI:3799130
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 427)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     The Dictyostelium developmental cDNA project: generation and
            analysis of expressed sequence tags from the first-finger stage of
            development
  JOURNAL   DNA Res. 5 (6), 335-340 (1998)
   PUBMED   10048482
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp
            PROJECT = 'Dictyostelium discoideum cDNA project in Japan'
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .427
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SLB340"
                     /dev_stage="slug"
                     /clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
ORIGIN

Alignment Scores:
Pred. No.:                2.45e+03       Length:         427
Score:                    42.00          Matches:        6
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       1              Gaps:           0

US-09-858-332G-28 (1-6) x AU033706 (1-427)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        207 CACAATCACAATCACAAT 224

RESULT 318
AU033893
LOCUS       AU033893                  427 bp    mRNA    linear   EST 28-APR-1999
DEFINITION  AU033893 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
            discoideum cDNA clone SLB578, mRNA sequence.
ACCESSION   AU033893
VERSION     AU033893.1  GI:3799317
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 427)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     The Dictyostelium developmental cDNA project: generation and
            analysis of expressed sequence tags from the first-finger stage of
            development
  JOURNAL   DNA Res. 5 (6), 335-340 (1998)
   PUBMED   10048482
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp
            PROJECT = 'Dictyostelium discoideum cDNA project in Japan'
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .427
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SLB578"
                     /dev_stage="slug"
                     /clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
ORIGIN

Alignment Scores:
Pred. No.:                2.45e+03       Length:         427
Score:                    42.00          Matches:        6
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       1              Gaps:           0

US-09-858-332G-28 (1-6) x AU033893 (1-427)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        207 CACAATCACAATCACAAT 224

RESULT 319
AV416079
LOCUS       AV416079                  427 bp    mRNA    linear   EST 10-MAR-2004
DEFINITION  AV416079 Lotus japonicus young plants (two-week old) Lotus
            corniculatus var. japonicus cDNA clone MWM121a11_r 5', mRNA
            sequence.
ACCESSION   AV416079
VERSION     AV416079.1  GI:7745259
```

```
KEYWORDS    EST.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1  (bases 1 to 427)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Generation of 7137 non-redundant expressed sequence tags from a
            legume, Lotus japonicus
  JOURNAL   DNA Res. 7 (2), 127-130 (2000)
   PUBMED   10819328
COMMENT     Contact: Yasukazu Nakamura
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: ynakamu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .427
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="mRNA"
                     /isolate="Miyakojima MG-20"
                     /db_xref="taxon:34305"
                     /clone="MWM121a11_r"
                     /tissue_type="whole plant"
                     /dev_stage="two-week old plant"
                     /clone_lib="Lotus japonicus young plants (two-week old)"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI; synonym: Lotus japonicus"
ORIGIN

Alignment Scores:
Pred. No.:                2.45e+03        Length:         427
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       1               Gaps:           0

US-09-858-332G-28 (1-6) x AV416079 (1-427)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db          9 CACAACCACAACCACAAC 26

RESULT 320
AV596087/c
LOCUS       AV596087                427 bp    mRNA    linear   EST 22-FEB-2005
DEFINITION  AV596087 Bos taurus cartilage fetus Bos taurus cDNA clone
            E1CA023F02 5', mRNA sequence.
ACCESSION   AV596087
VERSION     AV596087.1  GI:9713113
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 427)
  AUTHORS   Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.
            and Sugimoto,Y.
  TITLE     Establishment of a high throughput EST sequencing system using
            poly(A) tail-removed cDNA libraries and determination of 36,000
            bovine ESTs
  JOURNAL   Nucleic Acids Res. 29 (22), E108 (2001)
   PUBMED   11713328
COMMENT     Contact: Yoshikazu Sugimoto
            Animal Genetics Division
            Shirakawa Institute of Animal Genetics
            Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
            Tel: 81-248-25-5641
            Fax: 81-248-25-5725
            Email: kazusugi@siag.or.jp
            Single pass sequencing.
            This clone was obtained from a polyA-deleted cDNA library.
FEATURES             Location/Qualifiers
     source          1. .427
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /clone="E1CA023F02"
                     /tissue_type="cartilage"
                     /dev_stage="fetus"
                     /lab_host="DH10B"
                     /clone_lib="Bos taurus cartilage fetus"
                     /note="Vector: pZL1; Site_1: Sal1; Site_2: Not1; Poly A
                     was deleted from a Not1 site"
ORIGIN

Alignment Scores:
Pred. No.:                2.45e+03        Length:         427
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       1               Gaps:           0

US-09-858-332G-28 (1-6) x AV596087 (1-427)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        273 CATAACCACAACCATAAC 256

RESULT 321
BP805483/c
LOCUS       BP805483                427 bp    mRNA    linear   EST 15-FEB-2005
DEFINITION  BP805483 RAFL16 Arabidopsis thaliana cDNA clone RAFL24-06-C01 5',
            mRNA sequence.
ACCESSION   BP805483
VERSION     BP805483.1  GI:59841562
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 427)
  AUTHORS   Seki,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M., Akiyama,K.,
            Iida,K., Enju,A., Sakurai,T., Arakawa,T., Carninci,P., Fukuda,S.,
            Iida,J., Kawai,J., Sasaki,D., Shiraki,T., Hayashizaki,Y. and
            Shinozaki,K.
  TITLE     Large-Scale Analysis of RIKEN Arabidopsis Full-length cDNAs
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            An Arabidopsis full-length cDNA library was constructed essentially
            as reported previously (Seki et al., 1998, 2002).This clone is in a
            modified pBluescript vector.
            Please visit our web site (http://pfgweb.gsc.riken.jp and
            http://rarge.gsc.riken.jp) for further details.
FEATURES             Location/Qualifiers
     source          1. .427
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="RAFL24-06-C01"
                     /lab_host="DH10B"
                     /clone_lib="RAFL16"
                     /note="Site_1: BamHI; Site_2: SalI; dark-grown"
ORIGIN
```

```
Alignment Scores:
Pred. No.:                2.45e+03        Length:          427
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       3               Gaps:            0

US-09-858-332G-28 (1-6) x BP805483 (1-427)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        184 CACAATCACAATCACAAT 167

RESULT 322
CF542950
LOCUS       CF542950                 427 bp    mRNA    linear   EST 22-SEP-2003
DEFINITION  S014678-024-030-K14-SP6 MPIZ-ADIS-024-leaf Beta vulgaris cDNA clone
            024-030-K14 5-PRIME, mRNA sequence.
ACCESSION   CF542950
VERSION     CF542950.1  GI:34891390
KEYWORDS    EST.
SOURCE      Beta vulgaris
  ORGANISM  Beta vulgaris
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            Caryophyllales; Amaranthaceae; Beta.
REFERENCE   1  (bases 1 to 427)
  AUTHORS   Herwig,R., Schulz,B., Weisshaar,B., Hennig,S., Steinfath,M.,
            Drungowski,M., Stahl,D., Wruck,W., Menze,A., O'Brien,J., Lehrach,H.
            and Radelof,U.
  TITLE     Construction of a 'unigene' cDNA clone set by oligonucleotide
            fingerprinting allows access to 25 000 potential sugar beet genes
  JOURNAL   Plant J. 32 (5), 845-857 (2002)
   PUBMED   12472698
COMMENT     Contact: Weisshaar B
            ADIS DNA core facility at MPIZ
            Max-Planck-Institute for Plant Breeding Research
            Carl-von-Linne Weg 10, 50829 Koeln, Germany
            Fax: 00492215062851
            Email: weisshaa@mpiz-koeln.mpg.de
            Insert Length: 427   Std Error: 0.00
            Plate: 30  row: K  column: 14
            Seq primer: SP6.
FEATURES             Location/Qualifiers
     source          1. .427
                     /organism="Beta vulgaris"
                     /mol_type="mRNA"
                     /cultivar="KWS2320 (double haploid, monogerm breeding
                     line)"
                     /db_xref="GABI:936823"
                     /db_xref="taxon:161934"
                     /clone="024-030-K14"
                     /tissue_type="leaf"
                     /lab_host="EMDH10B"
                     /clone_lib="MPIZ-ADIS-024-leaf"
                     /note="Vector: pCMVSPORT6; Site_1: SalI; Site_2: NotI;
                     cDNA library from sugar beet, library provided by KWS
                     Kleinwanzlebener Saatzucht AG Einbeck, Germany, contact:
                     b.schulz@kws.de; cloning sites SalI-NotI, primer sites and
                     orientation:
                     SP6-SalI-CCACGCGTCCG-5prime-cDNA-polyA-CC-NotI-T7;  Note:
                     Sequencing granted in the context of the GABI-Beet
                     project, local PI: Dr. Katharina Schneider, coordinator:
                     Prof. Christian Jung;  Sequence submission managed by
                     RZPD/GABI-Primary database:http://gabi.rzpd.de"
ORIGIN

Alignment Scores:
Pred. No.:                2.45e+03        Length:          427
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       6               Gaps:            0

US-09-858-332G-28 (1-6) x CF542950 (1-427)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        110 CATAACCATAACCATAAC 127

RESULT 323
CK760961
LOCUS       CK760961                 427 bp    mRNA    linear   EST 09-JUN-2005
DEFINITION  nad03-11cs2-d05 Nad03 Nuphar advena cDNA clone nad03-11cs2-d05 5',
            mRNA sequence.
ACCESSION   CK760961
VERSION     CK760961.1  GI:42651402
KEYWORDS    EST.
SOURCE      Nuphar advena
  ORGANISM  Nuphar advena
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; basal Magnoliophyta; Nymphaeales;
            Nymphaeaceae; Nuphar.
REFERENCE   1  (bases 1 to 427)
  AUTHORS   dePamphilis,C., Carlson,J., Ma,H., Frohlich,M., Tanksley,S.,
            Leebens-Mack,J., Field,D., Arrington,J., Zahn,L., Kong,H.,
            Druckenmiller,M., Landherr,L., Hu,Y., Ilut,D., Wall,K.,
            Plock,S.Chiorean,S., Albert,V., Doyle,J., Miller,W.,
            Oppenheimer,D., Soltis,D., Soltis,P. and Theissen,G.
  TITLE     Generation of ESTs from early flower buds of Nuphar advena
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Claude dePamphilis or James Leebens-Mack
            Mueller Laboratory
            Penn State University
            208 Mueller Laboratory, Department of Biology, ATTN Rm212, Penn
            State University, University Park, PA 16802, USA
            Tel: 814 863 6413
            Fax: 814 865 9131
            Email: cwd3@psu.edu or jhl10@psu.edu
            The sequence provided is trimmed of vector and low quality regions.
            Full sequence and original trace file are available from the Plant
            Genome Network website (http://pgn.cornell.edu)
            Plate: nad03-11cs2  row: d  column: 05
            Seq primer: M13F.
FEATURES             Location/Qualifiers
     source          1. .427
                     /organism="Nuphar advena"
                     /mol_type="mRNA"
                     /db_xref="PGN:nad03-11cs2-d05"
                     /db_xref="taxon:77108"
                     /clone="nad03-11cs2-d05"
                     /tissue_type="flower buds"
                     /dev_stage="<= 2.5mm"
                     /lab_host="SOLR"
                     /clone_lib="Nad03"
                     /note="Vector: pBluescript SK (+/-); Site_1: EcoRI;
                     Site_2: XhoI; Only floral buds with diameter of 2.5 mm of
                     less were used for RNA isolation. This is a directionally
                     cloned, non-normalized library. Avg. insert length: 1134;
                     Primers: M13F and M13R; Antibiotic: 50 ug/ml Ampicillin;
                     Primary Titer: 2E6 pfu total; Amplified Titer: 3.2E10
                     pfu/ml; Mass Excised Titer: 5E10 total; This library has
                     been generated by the Floral Genome Project (FGP). The
                     Floral Genome Project is funded by NSF's Plant Genome
                     Research Program (DBI-0115684). More information about the
                     project can be obtained at http://fgp.bio.psu.edu"
ORIGIN

Alignment Scores:
Pred. No.:                2.45e+03        Length:          427
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
```

```
Query Match:            100.00%       Indels:        0
DB:                     7             Gaps:          0

US-09-858-332G-28 (1-6) x CK760961 (1-427)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         51 CATAATCATAATCATAAT 68

RESULT 324
CN729641/c
LOCUS       CN729641                427 bp    mRNA    linear   EST 14-MAY-2005
DEFINITION  EX20LIB3_UP_001_A11_16DEC2003_095 Brassica napus EX20LIB3 Brassica
            napus cDNA 5', mRNA sequence.
ACCESSION   CN729641
VERSION     CN729641.1  GI:65287445
KEYWORDS    EST.
SOURCE      Brassica napus (rape)
  ORGANISM  Brassica napus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 427)
  AUTHORS   Georges,F.
  TITLE     Gene expression at early stages of B. napus seed development as
            revealed by transcript profiling under normal and abiotic stress
            conditions
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Fawzy Georges
            Plant Biotechnology Institute, National Research Council Canada
            110 Gymnasium Place, Saskatoon, Saskatchewan, Canada S7N 7R9
            Tel: 306 975 4839
            Fax: 306 975 4815
            Email: Fawzy.Georges@nrc-cnrc.gc.ca
            High quality sequence stop: 427.
FEATURES             Location/Qualifiers
     source          1. .427
                     /organism="Brassica napus"
                     /mol_type="mRNA"
                     /db_xref="taxon:3708"
                     /clone_lib="Brassica napus EX20LIB3"
                     /note="Vector: pGemT-Easy; Sequences have been trimmed to
                     remove vector and low quality regions using LUCY sequence
                     cleanup software (www.tigr.org)."
ORIGIN

Alignment Scores:
Pred. No.:              2.45e+03      Length:        427
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     7             Gaps:          0

US-09-858-332G-28 (1-6) x CN729641 (1-427)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         33 CATAACCATAACCATAAC 16

RESULT 325
DR699194
LOCUS       DR699194                427 bp    mRNA    linear   EST 14-JUL-2005
DEFINITION  An_3235.A02 Aspergillus niger pYES2 (XhoI-EcoRI) Aspergillus niger
            cDNA clone An_3235, mRNA sequence.
ACCESSION   DR699194
VERSION     DR699194.1  GI:70815485
KEYWORDS    EST.
SOURCE      Aspergillus niger
  ORGANISM  Aspergillus niger
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE   1  (bases 1 to 427)
  AUTHORS   Tsang,A., Storms,R. and Bulter,G.
  TITLE     Expressed sequence tags from Aspergillus niger cDNA library
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Tsang A
            Centre for Structural and Functional Genomics
            Concordia University
            7141 Sherbrooke Street West, SP545-1, Montreal, QC H4B1R6, Canada
            Tel: 514 848 2424 3405
            Fax: 514 848 4504
            Email: tsang@vax2.concordia.ca
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .427
                     /organism="Aspergillus niger"
                     /mol_type="mRNA"
                     /strain="N400"
                     /db_xref="taxon:5061"
                     /clone="An_3235"
                     /dev_stage="mycelial growth"
                     /lab_host="E. coli"
                     /clone_lib="Aspergillus niger pYES2 (XhoI-EcoRI)"
                     /note="Vector: YES2; Site_1: XhoI; Site_2: EcoRI;
                     Complementary DNA was synthesized with ZAP kit
                     (Stratagene) using poly(A)+RNA isolated from Aspergillus
                     niger grown in both complete and minimal media. Synthesis
                     was primed with oligo(dT)/XhoI primer. EcoRI adaptors were
                     ligated to the blunt-ended, double-stranded cDNA.
                     EcoRI-XhoI-digested cDNA was ligated with
                     EcoRI-XhoI-digested pYES2 (Invitrogen Corp). This vector
                     permits expression of our library in yeast."
ORIGIN

Alignment Scores:
Pred. No.:              2.45e+03      Length:        427
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     8             Gaps:          0

US-09-858-332G-28 (1-6) x DR699194 (1-427)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        140 CACAACCACAACCACAAC 157

RESULT 326
H06653
LOCUS       H06653                  427 bp    mRNA    linear   EST 21-JUN-1995
DEFINITION  yl82b11.s1 Soares infant brain 1NIB Homo sapiens cDNA clone
            IMAGE:44597 3', mRNA sequence.
ACCESSION   H06653
VERSION     H06653.1  GI:870185
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 427)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
```

Insert Size: 1418
High quality sequence stops: 297
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1418 Std Error: 0.00
Seq primer: Promega -21m13
High quality sequence stop: 297.

FEATURES Location/Qualifiers
source 1. .427
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:417138"
/db_xref="taxon:9606"
/clone="IMAGE:44597"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares infant brain 1NIB"
/note="Organ: whole brain; Vector: Lafmid BA; Site_1: Not I; Site_2: Hind III; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Hind III adaptors (Pharmacia), digested with Not I and directionally cloned into the Not I and Hind III sites of the Lafmid BA vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

ORIGIN

Alignment Scores:

Pred. No.:	2.45e+03	Length:	427
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	8	Gaps:	0

US-09-858-332G-28 (1-6) x H06653 (1-427)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        208 CATAATCATAATCATAAT 225
```

RESULT 327
BG604753
LOCUS BG604753 428 bp mRNA linear EST 16-APR-2001
DEFINITION WHE0941_G06_M11ZS Wheat 5-15 DAP spike cDNA library Triticum aestivum cDNA clone WHE0941_G06_M11, mRNA sequence.
ACCESSION BG604753
VERSION BG604753.1 GI:13634756
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 428)
AUTHORS Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L. and Tong,J.C.
TITLE The structure and function of the expressed portion of the wheat genomes - 5-15 DAP spike cDNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oandersn@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20

Seq primer: Stratagene SK primer.

FEATURES Location/Qualifiers
source 1. .428
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE0941_G06_M11"
/tissue_type="Spike"
/dev_stage="Adult plant"
/lab_host="E. coli SOLR"
/clone_lib="Wheat 5-15 DAP spike cDNA library"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site_1: EcoRI; Site_2: XhoI; Plants were grown in the greenhouse. Spikes at 5, 10 and 15 DAP were harvested, total RNA and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give pBluescript phagemids in the TJ Close lab (Choi, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

ORIGIN

Alignment Scores:

Pred. No.:	2.45e+03	Length:	428
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-28 (1-6) x BG604753 (1-428)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         67 CACAACCACAACCACAAT 84
```

RESULT 328
AV421073
LOCUS AV421073 429 bp mRNA linear EST 10-MAR-2004
DEFINITION AV421073 Lotus japonicus young plants (two-week old) Lotus corniculatus var. japonicus cDNA clone MWM190a09_r 5', mRNA sequence.
ACCESSION AV421073
VERSION AV421073.1 GI:7774322
KEYWORDS EST.
SOURCE Lotus corniculatus var. japonicus (Lotus japonicus)
ORGANISM Lotus corniculatus var. japonicus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae; Lotus.
REFERENCE 1 (bases 1 to 429)
AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
TITLE Generation of 7137 non-redundant expressed sequence tags from a legume, Lotus japonicus
JOURNAL DNA Res. 7 (2), 127-130 (2000)
PUBMED 10819328
COMMENT Contact: Yasukazu Nakamura
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: ynakamu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.

FEATURES Location/Qualifiers
source 1. .429
/organism="Lotus corniculatus var. japonicus"
/mol_type="mRNA"
/isolate="Miyakojima MG-20"
/db_xref="taxon:34305"
/clone="MWM190a09_r"
/tissue_type="whole plant"
/dev_stage="two-week old plant"

```
                /clone_lib="Lotus japonicus young plants (two-week old)"
                /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                XhoI; synonym: Lotus japonicus"
ORIGIN

Alignment Scores:
Pred. No.:              2.46e+03        Length:         429
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AV421073 (1-429)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db          7 CACAACCACAACCACAAC 24

RESULT 329
BY034160
LOCUS       BY034160                429 bp    mRNA    linear   EST 06-DEC-2002
DEFINITION  BY034160 RIKEN full-length enriched, pooled tissues, 14 days
            embryo, etc. Mus musculus cDNA clone I530009N19 5', mRNA sequence.
ACCESSION   BY034160
VERSION     BY034160.1  GI:26139603
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 429)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
            Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
            Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
            Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
            Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
            Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
            Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
            Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
            Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
            Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
            Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
            Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
            Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
            Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
            Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
            Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
            Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
            Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
            Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
            Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
            Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
            Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
            Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
            Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
            Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
            Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
   PUBMED   12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
            Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
            Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,
            Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,
            Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and
            Hayashizaki,Y. Direct Submission
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
             cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES            Location/Qualifiers
     source         1. .429
                    /organism="Mus musculus"
                    /mol_type="mRNA"
                    /db_xref="taxon:10090"
                    /clone="I530009N19"
                    /clone_lib="RIKEN full-length enriched, pooled tissues, 14
                    days embryo, etc."
                    /note="pooled tissues: (dev_stage=14 days
                    embryo,tissue_type=liver,sex=mix),(dev_stage=11 days
                    pregnant, adult,tissue_type=placenta,sex=female),
                    (dev_stage=12 days pregnant,
                    adult,tissue_type=placenta,sex=female),(dev_stage=13 days
                    pregnant, adult,tissue_type=placenta,sex=female),
                    (dev_stage=14 days pregnant,
                    adult,tissue_type=placenta,sex=female),(dev_stage=15 days
                    pregnant, adult,tissue_type=placenta,sex=female),
                    (dev_stage=14 days pregnant,
                    adult,tissue_type=amnion,sex=female)"
ORIGIN

Alignment Scores:
Pred. No.:              2.46e+03        Length:         429
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     5               Gaps:           0

US-09-858-332G-28 (1-6) x BY034160 (1-429)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        303 CACAACCATAATCACAAC 320

RESULT 330
AI152323
LOCUS       AI152323                430 bp    mRNA    linear   EST 30-SEP-1998
DEFINITION .ud79a02.r1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
            IMAGE:1477034 5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0
            KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION   AI152323
VERSION     AI152323.1  GI:3680792
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 430)
```

```
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:925390
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 419.
FEATURES              Location/Qualifiers
     source           1. .430
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:1477034"
                      /sex="female (lactating)"
                      /tissue_type="mammary gland"
                      /lab_host="DH10B"
                      /clone_lib="Soares_mammary_gland_NMLMG"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; 1st strand cDNA was prepared from mammary
                      gland tissue from a lactating female, and was then primed
                      with a Not I - oligo(dT) primer. Double-stranded  cDNA was
                      ligated to Eco RI adaptors (Pharmacia), digested with Not
                      I  and cloned into the Not I and Eco RI sites of the
                      modified pT7T3 vector.  Library is normalized.  Library
                      was constructed by Bento Soares and M. Fatima Bonaldo. "
ORIGIN

Alignment Scores:
Pred. No.:                2.46e+03        Length:          430
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       1               Gaps:            0

US-09-858-332G-28 (1-6) x AI152323 (1-430)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        174 CACAACCATAACCACAAC 191

RESULT 331
BE604994
LOCUS       BE604994                 430 bp    mRNA    linear   EST 21-AUG-2000
DEFINITION  WHE1713-1716_E15_E15ZS Wheat heat stressed spike cDNA library
            Triticum aestivum cDNA clone WHE1713-1716_E15_E15, mRNA sequence.
ACCESSION   BE604994
VERSION     BE604994.1  GI:9862264
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 430)
  AUTHORS   Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D.,
            Han,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Malatrasi,M., Miller,R.,
            Nguyen,H.T., Rausch,C.J., Seaton,C.L., Tong,J.C. and Zhang,D.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Heat stressed spike cDNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene SK primer.
FEATURES              Location/Qualifiers
     source           1. .430
                      /organism="Triticum aestivum"
                      /mol_type="mRNA"
                      /cultivar="Chinese Spring"
                      /db_xref="taxon:4565"
                      /clone="WHE1713-1716_E15_E15"
                      /tissue_type="Whole spike"
                      /dev_stage="Spikes at 5, 10, 15 and 20 days after
                      anthesis"
                      /lab_host="E. coli SOLR"
                      /clone_lib="Wheat heat stressed spike cDNA library"
                      /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                      Site_1: EcoRI; Site_2: XhoI; Spikes at 5, 10, 15 and 20
                      days after anthesis were heat stressed under  two
                      conditions at Texas Tech University (D. Zhang in HT Nguyrn
                      lab):  (1) at 38 C for 4 hours and (2) 5 days of cyclic
                      treatment of 38 C  for 4 hours.  Total RNA and poly(A) RNA
                      were prepared, a cDNA library  was made, and the cDNA
                      clones were in vivo excised to give pBluescript  phagemids
                      in the TJ Close lab (Choi, Close, Fenton, Malatrasi) at
                      the  University of California, Riverside.  Plasmid DNA
                      preparations and  DNA sequencing were performed in the OD
                      Anderson lab (all other authors)."
ORIGIN

Alignment Scores:
Pred. No.:                2.46e+03        Length:          430
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       2               Gaps:            0

US-09-858-332G-28 (1-6) x BE604994 (1-430)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        277 CACAACCACAACCACAAT 294

RESULT 332
BF012441
LOCUS       BF012441                 430 bp    mRNA    linear   EST 06-OCT-2000
DEFINITION  ux56d02.y1 Soares_NKWMD_mandible Mus musculus cDNA clone
            IMAGE:3514275 5' similar to TR:Q9WTZ9 Q9WTZ9 BRAIN EXPRESSED
            X-LINKED PROTEIN 3. ;, mRNA sequence.
ACCESSION   BF012441
VERSION     BF012441.1  GI:10712716
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 430)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Other_ESTs: ux56d02.x1
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
```

```
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1397155
            Seq primer: -40RP from Gibco
            High quality sequence stop: 419.
FEATURES              Location/Qualifiers
     source           1. .430
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:3514275"
                      /tissue_type="mandible"
                      /lab_host="DH10B (phage-resistant)"
                      /clone_lib="Soares_NKWMD_mandible"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: NotI; Site_2: EcoRI; 1st strand cDNA
                      was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCCTTAATTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT7T3 vector. Library
                      went through one round of normalization. Library
                      constructed by Bento Soares and M. Fatima Bonaldo. "
ORIGIN

Alignment Scores:
Pred. No.:                2.46e+03        Length:          430
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       2               Gaps:            0

US-09-858-332G-28 (1-6) x BF012441 (1-430)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        162 CACAACCATAACCACAAC 179

RESULT 333
BU017925/c
LOCUS       BU017925                430 bp    mRNA    linear   EST 23-AUG-2002
DEFINITION  QHE17D20.yg.ab1 QH_EFGHJ sunflower RHA280 Helianthus annuus cDNA
            clone QHE17D20, mRNA sequence.
ACCESSION   BU017925
VERSION     BU017925.1  GI:22453445
KEYWORDS    EST.
SOURCE      Helianthus annuus (common sunflower)
  ORGANISM  Helianthus annuus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; campanulids; Asterales; Asteraceae; Asteroideae;
            Heliantheae; Helianthus.
REFERENCE   1  (bases 1 to 430)
  AUTHORS   Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J.,
            Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y.,
            Lai,Z., Church,S., Jackson,L. and Bradford,K.
  TITLE     Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
            singleton, see http://cgpdb.ucdavis.edu/ for details.
            Plate: QHE17  row: D  column: 20.
FEATURES              Location/Qualifiers
     source           1. .430
                      /organism="Helianthus annuus"
                      /mol_type="mRNA"
                      /cultivar="RHA280"
                      /db_xref="taxon:4232"
                      /clone="QHE17D20"
                      /lab_host="E.coli"
                      /clone_lib="QH_EFGHJ sunflower RHA280"
                      /note="Vector: pBRcDNASfiAB; The library was constructed
                      from 11 different sources of RNA from a single genotype.
                      Separate cDNAs were generated using primers that
                      incorporated unique 5' and 3' tags to distinguish each
                      source of RNA. cDNAs were then pooled, size-fractionated,
                      directionally cloned into a custom medium-copy vector and
                      transformations made with four size classes to minimize
                      size bias. Details of each source of RNA and library
                      construction can be obtained at http://cgpdb.ucdavis.edu/
                      TAG_TISSUE=chemical induction
                      TAG_LIB=QH_EFGHJ sunflower RHA280
                      TAG_SEQ=TGTAGCCGGG"
ORIGIN

Alignment Scores:
Pred. No.:                2.46e+03        Length:          430
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       5               Gaps:            0

US-09-858-332G-28 (1-6) x BU017925 (1-430)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         99 CATAACCATAACCATAAC 82

RESULT 334
CF987483
LOCUS       CF987483                430 bp    mRNA    linear   EST 04-FEB-2005
DEFINITION  15498rsicee_11376.y1 Oryza sativa cv. LYP9 tillering whole plant
            cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF987483
VERSION     CF987483.1  GI:58589175
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 430)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 430
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .430
                      /organism="Oryza sativa (indica cultivar-group)"
                      /mol_type="mRNA"
                      /cultivar="LYP9"
```

```
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="tillering"
                     /clone_lib="Oryza sativa cv. LYP9 tillering whole plant
                     cDNA library"
ORIGIN

Alignment Scores:
Pred. No.:                2.46e+03        Length:         430
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       7               Gaps:           0

US-09-858-332G-28 (1-6) x CF987483 (1-430)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        217 CACAACCACAACCACAAC 234

RESULT 335
CO320011/c
LOCUS       CO320011                  430 bp    mRNA    linear   EST 28-JUN-2004
DEFINITION  EK281720.5prime Exelixis FlyTag CK01 pCDNA-SK+ Drosophila
            melanogaster cDNA clone EK281720 5, mRNA sequence.
ACCESSION   CO320011
VERSION     CO320011.1  GI:49378445
KEYWORDS    EST.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 430)
  AUTHORS   Kopczynski,C., Platt,D., Campbell,J., Muzong,C., Laufer,A.,
            Peterson,E. and Swimmer,C.
  TITLE     Exelixis FlyTag EST Project CK01 Library
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: EK.2817  row: B  column: 8
            High quality sequence stop: 361.
FEATURES             Location/Qualifiers
     source          1. .430
                     /organism="Drosophila melanogaster"
                     /mol_type="mRNA"
                     /db_xref="taxon:7227"
                     /clone="EK281720"
                     /clone_lib="Exelixis FlyTag CK01 pCDNA-SK+"
                     /note="Organ: mixed stage embryos, imaginal disks, and
                     adult heads; Vector: pCDNA-SK+; Site_1: NotI; Site_2:
                     XhoI; Random primed, normalized library from mixed stage
                     embryos, imaginal disks, and adult heads."
ORIGIN

Alignment Scores:
Pred. No.:                2.46e+03        Length:         430
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       7               Gaps:           0

US-09-858-332G-28 (1-6) x CO320011 (1-430)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        194 CATAATCATAATCATAAT 177

RESULT 336
BG071356/c
LOCUS       BG071356                  431 bp    mRNA    linear   EST 13-JUN-2003
DEFINITION  H3097H01-3 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
            H3097H01 3', mRNA sequence.
ACCESSION   BG071356
VERSION     BG071356.2  GI:31742756
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 431)
  AUTHORS   Tanaka,T.S., Jaradat,S.A., Lim,M.K., Kargul,G.J., Wang,X.,
            Grahovac,M.J., Pantano,S., Sano,Y., Piao,Y., Nagaraja,R., Doi,H.,
            Wood,W.H. III, Becker,K.G. and Ko,M.S.H.
  TITLE     Genome-wide expression profiling of mid-gestation placenta and
            embryo using a 15,000 mouse developmental cDNA microarray
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)
  PUBMED    10922068
COMMENT     On Jan 26, 2001 this sequence version replaced gi:12553925.
            Other_ESTs: H3097H01-5
            Contact: George J. Kargul
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgsun.grc.nia.nih.gov
            This clone set has been freely distributed to the community. Please
            visit http://lgsun.grc.nia.nih.gov/cDNA/15k.html for details.
            Plate: H3097  row: H  column: 01
            Seq primer: -21M13 Forward
            High quality sequence stop: 431
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .431
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="niaEST:H3097H01-3"
                     /db_xref="taxon:10090"
                     /clone="H3097H01"
                     /sex="Clones arrayed from a variety of cDNA libraries"
                     /dev_stage="Clones arrayed from a variety of cDNA
                     libraries"
                     /lab_host="DH10B"
                     /clone_lib="NIA Mouse 15K cDNA Clone Set"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
                     clone is among a rearrayed set of 15,247 clones from 11
                     embryo  cDNA libraries (including preimplantation stage
                     embryos from  unfertilized egg to blastocyst, embryonic
                     part of E7.5 embryos,  extraembryonic part of E7.5
                     embryos, and E12.5 female mesonephros/gonad)  and one
                     newborn ovary cDNA library. Average insert size 1.5 kb.
                     All  source libraries are cloned unidirectionally with
                     Oligo(dT)-Not primers.  References include: (1)
                     Genome-wide expression profiling of mid-gestation
                     placenta and embryo using a 15,000 mouse developmental
                     cDNA microarray,  2000, Proc. Natl. Acad. Sci. U S A, 97:
                     9127-9132; (2) Large-scale cDNA  analysis reveals phased
                     gene expression patterns during preimplantation  mouse
                     develolpment, 2000, Development, 127: 1737-1749; (3)
                     Genome-wide  mapping of unselected transcripts from
                     extraembryonic tissue of 7.5-day  mouse embryos reveals
                     enrichment in the t-complex and under-representation  on
                     the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."
ORIGIN

Alignment Scores:
Pred. No.:                2.47e+03        Length:         431
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
```

```
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     2              Gaps:          0

US-09-858-332G-28 (1-6) x BG071356 (1-431)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        395 CACAACCACAACCACAAC 378

RESULT 337
CA715331
LOCUS       CA715331                 431 bp    mRNA    linear   EST 26-NOV-2002
DEFINITION  wdk3c.pk012.k15 wdk3c Triticum aestivum cDNA clone wdk3c.pk012.k15
            5' end, mRNA sequence.
ACCESSION   CA715331
VERSION     CA715331.1  GI:25437124
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 431)
  AUTHORS   Tingey,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z.,
            Miao,G., Caraher,N. and Hanafey,M.K.
  TITLE     DuPont Wheat cDNA Sequence
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Scott V. Tingey
            Crop Genetics
            E. I. DuPont de Nemours and Company
            1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
            Tel: 302-631-2602
            Fax: 302-631-2607
            Email: Scott.V.Tingey@USA.dupont.com
            Seq primer: M13.
FEATURES             Location/Qualifiers
     source          1. .431
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /db_xref="taxon:4565"
                     /clone="wdk3c.pk012.k15"
                     /tissue_type="kernel"
                     /lab_host="DH10B"
                     /clone_lib="wdk3c"
                     /note="Vector: pBluescript SK+; Site_1: EcoRI; Site_2:
                     XhoI; Wheat (Triticum aestivum L.) developing kernel, 14
                     days after anthesis."
ORIGIN

Alignment Scores:
Pred. No.:              2.47e+03       Length:        431
Score:                  42.00          Matches:       6
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     6              Gaps:          0

US-09-858-332G-28 (1-6) x CA715331 (1-431)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        116 CACAACCACAACCACAAC 133

RESULT 338
CO294672
LOCUS       CO294672                 431 bp    mRNA    linear   EST 25-JUN-2004
DEFINITION  EK207425.5prime Exelixis FlyTag CK01 pCDNA-SK+ Drosophila
            melanogaster cDNA clone EK207425 5, mRNA sequence.
ACCESSION   CO294672
VERSION     CO294672.1  GI:49216218
KEYWORDS    EST.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 431)
  AUTHORS   Kopczynski,C., Platt,D., Campbell,J., Muzong,C., Laufer,A.,
            Peterson,E. and Swimmer,C.
  TITLE     Exelixis FlyTag EST Project CK01 Library
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Based upon one or more reads of this clone where vector sequence
            was present at both ends, this clone has been determined to contain
            contain a cDNA insert on the order of 600-1000 bases.
            Plate: EK.2074  row: C  column: 1
            High quality sequence stop: 389.
FEATURES             Location/Qualifiers
     source          1. .431
                     /organism="Drosophila melanogaster"
                     /mol_type="mRNA"
                     /db_xref="taxon:7227"
                     /clone="EK207425"
                     /clone_lib="Exelixis FlyTag CK01 pCDNA-SK+"
                     /note="Organ: mixed stage embryos, imaginal disks, and
                     adult heads; Vector: pCDNA-SK+; Site_1: NotI; Site_2:
                     XhoI; Random primed, normalized library from mixed stage
                     embryos, imaginal disks, and adult heads."
ORIGIN

Alignment Scores:
Pred. No.:              2.47e+03       Length:        431
Score:                  42.00          Matches:       6
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     7              Gaps:          0

US-09-858-332G-28 (1-6) x CO294672 (1-431)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        214 CATAATCATAATCATAAT 231

RESULT 339
CV260941
LOCUS       CV260941                 432 bp    mRNA    linear   EST 22-SEP-2004
DEFINITION  WS02015.B21_G18 PTxN-IB-N-A-11 Populus trichocarpa x Populus nigra
            cDNA clone WS02015_G18 3', mRNA sequence.
ACCESSION   CV260941
VERSION     CV260941.1  GI:52513916
KEYWORDS    EST.
SOURCE      Populus trichocarpa x Populus nigra
  ORGANISM  Populus trichocarpa x Populus nigra
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Malpighiales; Salicaceae; Saliceae; Populus.
REFERENCE   1  (bases 1 to 432)
  AUTHORS   Ralph,S., Cooper,D., Kolosova,N., Oddy,C., Butterfield,Y.,
            Kirkpatrick,R., Liu,J., Palmquist,D., Stott,J., Barber,S., Yang,G.,
            Babakaiff,R., Brown-John,M., Chand,S., Featherstone,R., Masson,A.,
            Mayo,M., Moran,J., Olson,T., Wong,D., Ritland,C.E., Siddiqui,A.,
            Holt,R., Jones,S., Marra,M., Ellis,B.E., Douglas,C., Ritland,K. and
            Bohlmann,J.
  TITLE     The poplar transcriptome: Analysis of expressed sequence tags from
            multiple cDNA libraries
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Joerg Bohlmann
            Genome BC forest genomics program
            University of British Columbia
```

Michael Smith Laboratories, 6174 University Boulevard, Rm. 237,
Vancouver, British Columbia, Canada, V6T 1Z3
Tel: 1-604-822-0282
Fax: 1-604-822-2114
Email: bohlmann@msl.ubc.ca
Plate: WS02015 row: G column: 18
High quality sequence stop: 432
POLYA=Yes.

FEATURES Location/Qualifiers
source 1. .432
/organism="Populus trichocarpa x Populus nigra"
/mol_type="mRNA"
/cultivar="NxM6"
/db_xref="taxon:293756"
/clone="WS02015_G18"
/sex="Male"
/lab_host="E. coli DH10B T1 phage resistant cells"
/clone_lib="PTxN-IB-N-A-11"
/note="Vector: pBluescript II SK (+) XR; Site_1: EcoRI (5' end of cDNA); Site_2: XhoI (3' end of cDNA); Sapling trees two metres in height and grown under greenhouse conditions were exposed to continuous feeding by Cryptorhynchus lapathi (poplar and willow borer) adults caged on the sapling using mesh bags. Bark (with phloem and cambium attached) from within the caged region was harvested 2 hours, 6 hours and 48 hours after the onset of treatment. mRNA was isolated from each tissue source independently and equal quantities of mRNA from each tissue were then pooled. cDNA was prepared from 5 micrograms of mRNA and directionally ligated into the pBluescript II SK (+) XR vector using the pBluescript II XR cDNA Library Construction Kit according to manufacturer's instructions with modifications (Stratagene). Plasmid DNA was then transformed by electroporation into DH10B cells (Invitrogen) for propagation. Normalization was applied according to published methods [Bonaldo M.F. et al. (1996) Genome Research 6(9):791] in order to reduce the abundance of highly expressed transcripts."

ORIGIN

Alignment Scores:

Pred. No.:	2.48e+03	Length:	432
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	7	Gaps:	0

US-09-858-332G-28 (1-6) x CV260941 (1-432)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         69 CACAATCATAACCACAAC 86
```

RESULT 340
CV269305
LOCUS CV269305 432 bp mRNA linear EST 22-SEP-2004
DEFINITION WS0207.B21_O06 PTxN-IB-N-A-11 Populus trichocarpa x Populus nigra cDNA clone WS0207_O06 3', mRNA sequence.
ACCESSION CV269305
VERSION CV269305.1 GI:52522280
KEYWORDS EST.
SOURCE Populus trichocarpa x Populus nigra
ORGANISM Populus trichocarpa x Populus nigra
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids I; Malpighiales; Salicaceae; Saliceae; Populus.
REFERENCE 1 (bases 1 to 432)
AUTHORS Ralph,S., Cooper,D., Kolosova,N., Oddy,C., Butterfield,Y., Kirkpatrick,R., Liu,J., Palmquist,D., Stott,J., Barber,S., Yang,G., Babakaiff,R., Brown-John,M., Chand,S., Featherstone,R., Masson,A., Mayo,M., Moran,J., Olson,T., Wong,D., Ritland,C.E., Siddiqui,A., Holt,R., Jones,S., Marra,M., Ellis,B.E., Douglas,C., Ritland,K. and Bohlmann,J.
TITLE The poplar transcriptome: Analysis of expressed sequence tags from multiple cDNA libraries
JOURNAL Unpublished (2004)
COMMENT Contact: Joerg Bohlmann
Genome BC forest genomics program
University of British Columbia
Michael Smith Laboratories, 6174 University Boulevard, Rm. 237,
Vancouver, British Columbia, Canada, V6T 1Z3
Tel: 1-604-822-0282
Fax: 1-604-822-2114
Email: bohlmann@msl.ubc.ca
Plate: WS0207 row: O column: 06
High quality sequence stop: 432
POLYA=Yes.

FEATURES Location/Qualifiers
source 1. .432
/organism="Populus trichocarpa x Populus nigra"
/mol_type="mRNA"
/cultivar="NxM6"
/db_xref="taxon:293756"
/clone="WS0207_O06"
/sex="Male"
/lab_host="E. coli DH10B T1 phage resistant cells"
/clone_lib="PTxN-IB-N-A-11"
/note="Vector: pBluescript II SK (+) XR; Site_1: EcoRI (5' end of cDNA); Site_2: XhoI (3' end of cDNA); Sapling trees two metres in height and grown under greenhouse conditions were exposed to continuous feeding by Cryptorhynchus lapathi (poplar and willow borer) adults caged on the sapling using mesh bags. Bark (with phloem and cambium attached) from within the caged region was harvested 2 hours, 6 hours and 48 hours after the onset of treatment. mRNA was isolated from each tissue source independently and equal quantities of mRNA from each tissue were then pooled. cDNA was prepared from 5 micrograms of mRNA and directionally ligated into the pBluescript II SK (+) XR vector using the pBluescript II XR cDNA Library Construction Kit according to manufacturer's instructions with modifications (Stratagene). Plasmid DNA was then transformed by electroporation into DH10B cells (Invitrogen) for propagation. Normalization was applied according to published methods [Bonaldo M.F. et al. (1996) Genome Research 6(9):791] in order to reduce the abundance of highly expressed transcripts."

ORIGIN

Alignment Scores:

Pred. No.:	2.48e+03	Length:	432
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	7	Gaps:	0

US-09-858-332G-28 (1-6) x CV269305 (1-432)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         69 CACAATCATAACCACAAC 86
```

RESULT 341
CV475084
LOCUS CV475084 432 bp mRNA linear EST 01-OCT-2004
DEFINITION 23562.1 Developing Tubers Solanum tuberosum cDNA clone 23562 5', mRNA sequence.
ACCESSION CV475084
VERSION CV475084.1 GI:53699860
KEYWORDS EST.
SOURCE Solanum tuberosum (potato)
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;

```
            asterids; lamiids; Solanales; Solanaceae; Solanum.
REFERENCE   1  (bases 1 to 432)
  AUTHORS   Flinn,B., Rothwell,C., Sardana,R., Griffiths,R., Lague,M., De
            Koeyer,D., Audy,P., Goyer,C., Li,X-Q., Wang-Pruski,G. and Regan,S.
  TITLE     Generation of ESTs from a normalized developing potato tuber
            library
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Barry Flinn
            The Canadian Potato Genome Project - BioAtlantech
            921 College Hill Rd, Fredericton, ON, E3B 6Z9, CANADA
            Email: bflinn@bioatlantech.nb.ca
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .432
                     /organism="Solanum tuberosum"
                     /mol_type="mRNA"
                     /cultivar="Shepody"
                     /db_xref="taxon:4113"
                     /clone="23562"
                     /tissue_type="Tubers"
                     /lab_host="XL10-Gold"
                     /clone_lib="Developing Tubers"
                     /note="Vector: pBluescript II SK(+) XR; Site_1: EcoRI;
                     Site_2: XhoI; supplier: Developmental series. Plants from
                     pathogen-free Solanum tuberosum var. Shepody, clone 1756,
                     nuclear stock were grown in a screenhouse under natural
                     conditions. Developing tubers of 10-15 mm in length and
                     30-40 mm in length were collected early during the growing
                     season and pooled. Mature tubers were collected
                     approximately 1 3/4 months later at top kill, and then
                     collected two weeks later post-top kill and pooled. RNA
                     was isolated from the 2 pooled collections and mixed for
                     library construction. A normalized library was constructed
                     following a modified protocol of Bonaldo et al. (1996.
                     Genome Research 6: 791-806)."
ORIGIN

Alignment Scores:
Pred. No.:                2.48e+03       Length:          432
Score:                    42.00          Matches:         6
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       7              Gaps:            0

US-09-858-332G-28 (1-6) x CV475084 (1-432)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         12 CACAACCACAACCACAAC 29

RESULT 342
AJ611217
LOCUS       AJ611217                 433 bp    mRNA    linear   EST 11-DEC-2003
DEFINITION  AJ611217 Triticum turgidum subsp. durum etiolated seedling 20 day
            Triticum turgidum subsp. durum cDNA clone 04475R, mRNA sequence.
ACCESSION   AJ611217
VERSION     AJ611217.1  GI:39728144
KEYWORDS    EST.
SOURCE      Triticum turgidum subsp. durum (durum wheat)
  ORGANISM  Triticum turgidum subsp. durum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 433)
  AUTHORS   Cifarelli,R.A., D'Onofrio,O. and Lauria,G.
  TITLE     Expressed Sequences Tags (ESTs) library from totipotent cDNA of
            durum wheat
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Cifarelli RA
            Biotechnology
            Metapontum Agrobios
            S.S. Jonica 106 km 448.2, 75010 Metaponto (MT), Italy.
FEATURES             Location/Qualifiers
     source          1. .433
                     /organism="Triticum turgidum subsp. durum"
                     /mol_type="mRNA"
                     /cultivar="Ofanto"
                     /sub_species="durum"
                     /db_xref="taxon:4567"
                     /clone="04475R"
                     /tissue_type="etiolated seedling"
                     /dev_stage="20 day"
                     /clone_lib="Triticum turgidum subsp. durum etiolated
                     seedling 20 day"
ORIGIN

Alignment Scores:
Pred. No.:                2.48e+03       Length:          433
Score:                    42.00          Matches:         6
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       1              Gaps:            0

US-09-858-332G-28 (1-6) x AJ611217 (1-433)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        213 CACAACCACAACCACAAC 230

RESULT 343
N59345
LOCUS       N59345                   434 bp    mRNA    linear   EST 23-FEB-1996
DEFINITION  yz86f10.s1 Soares_multiple_sclerosis_2NbHMSP Homo sapiens cDNA
            clone IMAGE:289963 3', mRNA sequence.
ACCESSION   N59345
VERSION     N59345.1  GI:1203235
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 434)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: m13 -40 forward
            High quality sequence stop: 300.
FEATURES             Location/Qualifiers
     source          1. .434
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="GDB:3905803"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:289963"
                     /sex="male"
                     /tissue_type="multiple sclerosis lesions"
                     /dev_stage="Age 46"
                     /lab_host="DH10B (ampicillin resistant)"
                     /clone_lib="Soares_multiple_sclerosis_2NbHMSP"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker V_TYPE: phagemid; Site_1: Not I; Site_2: Eco
```

RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo. RNA from 4 multiple sclerosis lesions from one patient was kindly provided by Dr. Kevin G. Becker (NINDS/NIH). "

ORIGIN

Alignment Scores:

Pred. No.:	2.49e+03	Length:	434
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	8	Gaps:	0

US-09-858-332G-28 (1-6) x N59345 (1-434)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        191 CATAATCATAATCATAAT 208
```

RESULT 344
CL878923
LOCUS CL878923 435 bp DNA linear GSS 30-AUG-2004
DEFINITION abf27b10.x1 Soybean random, unfiltered genomic library Glycine max genomic, genomic survey sequence.
ACCESSION CL878923
VERSION CL878923.1 GI:51612241
KEYWORDS GSS.
SOURCE Glycine max (soybean)
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE 1 (bases 1 to 435)
AUTHORS Nunberg,A., Bedell,J.A., Citek,R.W., Robbins,D., McMenamy,J., Peterson,S., Jones,J., Fries,J., Budiman,M.A., Nguyen,H. and Stacey,G.
TITLE Methylation filtered genomic sequences from Glycine max
JOURNAL Unpublished (2004)
COMMENT Contact: Gary Stacey
University of Missouri
108 Waters Hall, Columbia, MO 65211, USA
Tel: 573-884-1267
Fax: 573-882-0588
Email: staceyg@missouri.edu
LidID: 230
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .435
/organism="Glycine max"
/mol_type="genomic DNA"
/cultivar="Williams 82"
/db_xref="taxon:3847"
/tissue_type="Young leaves"
/clone_lib="Soybean random, unfiltered genomic library"
/note="Vector: pOT2; Site_1: BstXI; Randomly sheared genomic DNA ranging from 0.7-1.5 kb were end repaired and ligated to BstXI linkers prior to cloning in BstXI-cut pOT2. LibID: 230"
ORIGIN

Alignment Scores:

Pred. No.:	2.5e+03	Length:	435
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	10	Gaps:	0

US-09-858-332G-28 (1-6) x CL878923 (1-435)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        305 CACAATCACAATCATAAT 322
```

RESULT 345
BM369212/c
LOCUS BM369212 436 bp mRNA linear EST 23-JUL-2002
DEFINITION EBem07_SQ002_O19_R embryo, 28 DPA, no treatment, cv Optic, EBem07 Hordeum vulgare subsp. vulgare cDNA clone EBem07_SQ002_O19 5', mRNA sequence.
ACCESSION BM369212
VERSION BM369212.2 GI:21936364
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 436)
AUTHORS Hedley,P., Liu,H., Caldwell,D., McCallum,N., Mudie,S., Cardle,L., Ramsay,L., Machray,G., Marshall,D.F.M. and Waugh,R.
TITLE Development of Barley Transcriptome Resources
JOURNAL Unpublished (2001)
COMMENT On Jan 10, 2002 this sequence version replaced gi:18112602.
Contact: Waugh R, Marshall DF
Genome Dynamics/Computational Biology
Scottish Crop Research Institute
Invergowrie, Dundee, DD2 5DA, Scotland, UK
Tel: 00 44 1382 562731
Fax: 00 44 1382 562426
Email: est@scri.sari.ac.uk
All sequence has a Phred quality score of 20 or over
Seq primer: M13 reverse.
FEATURES Location/Qualifiers
source 1. .436
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Optic"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="EBem07_SQ002_O19"
/tissue_type="embryo"
/dev_stage="28 DPA"
/lab_host="DH10B"
/clone_lib="embryo, 28 DPA, no treatment, cv Optic, EBem07"
/note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I; Non-normalised library, directionally cloned into pSPORT1. Derived from embryos dissected from developing grains (28 days post anthesis) in glasshouse grown barley plants. Developed as part of the barley transcriptome resources of BBSRC/SEERAD funded cereal IGF (Investigating Gene Function) project."
ORIGIN

Alignment Scores:

Pred. No.:	2.5e+03	Length:	436
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	3	Gaps:	0

US-09-858-332G-28 (1-6) x BM369212 (1-436)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
```

```
Db        172 CACAATCACAATCACAAT 155

RESULT 346
BF941012
LOCUS       BF941012                437 bp    mRNA    linear   EST 22-JAN-2001
DEFINITION  hu63b06.x1 NCI_CGAP_Brn41 Homo sapiens cDNA clone IMAGE:3174491 3',
            mRNA sequence.
ACCESSION   BF941012
VERSION     BF941012.1  GI:12358332
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 437)
  AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Robert Jenkins, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D. and M.Fatima
            Bonaldo, Ph.D
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer: -40UP from Gibco
            High quality sequence stop: 436.
FEATURES             Location/Qualifiers
     source          1. .437
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3174491"
                     /tissue_type="oligodendroglioma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NCI_CGAP_Brn41"
                     /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: NotI; Site_2: EcoRI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATCACTTTTTTTTTTTTTTTTT 3']
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not
                     I and Eco RI sites of the modified pT7T3 vector. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
ORIGIN

Alignment Scores:
Pred. No.:              2.51e+03        Length:         437
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-858-332G-28 (1-6) x BF941012 (1-437)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        195 CATAATCATAATCATAAT 212

RESULT 347
BJ234355
LOCUS       BJ234355                437 bp    mRNA    linear   EST 17-SEP-2002
DEFINITION  BJ234355 Y. Ogihara unpublished cDNA library, Wh_e Triticum
            aestivum cDNA clone whe9n01 5', mRNA sequence.
ACCESSION   BJ234355
VERSION     BJ234355.1  GI:23080205
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 437)
  AUTHORS   Ogihara,Y. and Murai,K.
  TITLE     Expressed genes in Triticum aestivum
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .437
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /cultivar="Chinese Spring"
                     /db_xref="taxon:4565"
                     /clone="whe9n01"
                     /tissue_type="seed DPA10"
                     /dev_stage="Feekes' scale 11.2"
                     /clone_lib="Y. Ogihara unpublished cDNA library, Wh_e"
ORIGIN

Alignment Scores:
Pred. No.:              2.51e+03        Length:         437
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-858-332G-28 (1-6) x BJ234355 (1-437)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        303 CACAACCACAACCACAAT 320

RESULT 348
CB964579/c
LOCUS       CB964579                437 bp    mRNA    linear   EST 29-APR-2003
DEFINITION  NL8_E02 Drought stress (leaf) Oryza sativa (indica cultivar-group)
            cDNA clone NL8_E02 3', mRNA sequence.
ACCESSION   CB964579
VERSION     CB964579.1  GI:30226687
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 437)
  AUTHORS   Markandeya,G., Ravindra Babu,P., Venkat Reddy,B., Nagabhushana,I.,
            Chandra Sekhar,A., Bennetzen,J.L., Ramakrishna,W. and Reddy,A.R.
  TITLE     ESTs from a normalized cDNA library of drought stressed rice
            seedlings (Oryza sativa L.cv Nagina 22)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Reddy AR
            Department of Plant Sciences, School of Life Sciences
            University of Hyderabad
            P.O. Central University, Hyderabad-500 046, A.P, India
            Tel: 0091-40-3010265
            Fax: 0091-40-3010145
            Email: arjulsl@uohyd.ernet.in
            Insert Length: 437   Std Error: 0.00
            Seq primer: CGCCAGGGTTTTCCCAGTCACGAC.
FEATURES             Location/Qualifiers
```

```
     source          1. .437
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nagina 22"
                     /db_xref="taxon:39946"
                     /clone="NL8_E02"
                     /tissue_type="Entire leaf tissue"
                     /dev_stage="35 day-old seedlings"
                     /clone_lib="Drought stress (leaf)"
                     /note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized
                     leaf cDNA Library from drought stressed seedlings "
ORIGIN

Alignment Scores:
Pred. No.:              2.51e+03        Length:         437
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x CB964579 (1-437)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        356 CACAACCACAACCACAAC 339

RESULT 349
CF965114
LOCUS       CF965114                437 bp    mRNA    linear   EST 15-JAN-2004
DEFINITION  9492rsiced_8168.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF965114
VERSION     CF965114.1  GI:38481025
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 437)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 437
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .437
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="trefoil"
                     /clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA
                     library"
ORIGIN

Alignment Scores:
Pred. No.:              2.51e+03        Length:         437
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     7               Gaps:           0

US-09-858-332G-28 (1-6) x CF965114 (1-437)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        194 CACAACCACAACCACAAC 211

RESULT 350
CO336724
LOCUS       CO336724                437 bp    mRNA    linear   EST 05-OCT-2004
DEFINITION  EN13121.5prime Exelixis FlyTag MN08 BlueScript Drosophila
            melanogaster cDNA clone EN13121 5, mRNA sequence.
ACCESSION   CO336724
VERSION     CO336724.1  GI:49396999
KEYWORDS    EST.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 437)
  AUTHORS   Nakanishi,M., Muzong,C., Peterson,E., Laufer,A., Leung,W., Platt,D.
            and Swimmer,C.
  TITLE     Exelixis FlyTag EST Project MN08 Library
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: EN.131  row: B  column: 9
            High quality sequence stop: 436.
FEATURES             Location/Qualifiers
     source          1. .437
                     /organism="Drosophila melanogaster"
                     /mol_type="mRNA"
                     /db_xref="taxon:7227"
                     /clone="EN13121"
                     /cell_line="mbn2"
                     /clone_lib="Exelixis FlyTag MN08 BlueScript"
                     /note="Vector: pBluescript; Site_1: NotI; Site_2: XhoI;
                     oligodT primed from LPS induced mbn2 cell line."
ORIGIN

Alignment Scores:
Pred. No.:              2.51e+03        Length:         437
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     7               Gaps:           0

US-09-858-332G-28 (1-6) x CO336724 (1-437)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        283 CATAATCACAACCATAAT 300

RESULT 351
AI488999
LOCUS       AI488999                438 bp    mRNA    linear   EST 18-MAY-2001
DEFINITION  EST247338 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
            cLED18O22, mRNA sequence.
ACCESSION   AI488999
```

```
VERSION     AI488999.1  GI:4384370
KEYWORDS    EST.
SOURCE      Lycopersicon esculentum (Solanum lycopersicum)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 438)
  AUTHORS   Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T.,
            Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B.,
            Fujii,C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C.,
            Martin,G.B., Tanksley,S.D. and Giovannoni,J.
  TITLE     Generation of ESTs from tomato carpel tissue
  JOURNAL   Unpublished (1999)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html.
FEATURES             Location/Qualifiers
     source          1. .438
                     /organism="Lycopersicon esculentum"
                     /mol_type="mRNA"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cLED18022"
                     /tissue_type="carpel"
                     /dev_stage="5 days pre-anthesis to 5 days post-anthesis"
                     /lab_host="XL1-Blue MRF'"
                     /clone_lib="tomato ovary, TAMU"
                     /note="Vector: pBlueScript SK(-); Site_1: EcoR1; Site_2:
                     Xho1; cLED - Tomato Carpel EST Library. OligodT-primed and
                     directionally cloned  cDNA in vector Lamda ZAP II with 5'
                     and 3' ends located at the EcoRI and  XhoI sites,
                     respectively."
ORIGIN

Alignment Scores:
Pred. No.:              2.51e+03      Length:         438
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     1             Gaps:           0

US-09-858-332G-28 (1-6) x AI488999 (1-438)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         43 CACAACCACAATCACAAC 60

RESULT 352
BE758972
LOCUS       BE758972                438 bp    mRNA    linear   EST 24-OCT-2000
DEFINITION  an_3235 Aspergillus niger, pYES2 (XhoI-EcoRI) Aspergillus niger
            cDNA clone 3235 5', mRNA sequence.
ACCESSION   BE758972
VERSION     BE758972.1  GI:10181609
KEYWORDS    EST.
SOURCE      Aspergillus niger
  ORGANISM  Aspergillus niger
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE   1  (bases 1 to 438)
  AUTHORS   Tsang,A. and Storms,R.
  TITLE     Aspergillus niger Expressed Sequence Tags
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Tsang, A.
            Dept. of Biology
            Concordia University
            1455 deMaisonneuve Blvd. West, Montreal, Quebec, CANADA, H3G 1M8
            Tel: 514-848-3405
            Fax: 514-848-4504
            Email: tsang@vax2.concordia.ca
            PCR PRimers
            FORWARD: CGACATTATGCTGAGTGATATCCC
            Insert Length: 438   Std Error: 0.00
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .438
                     /organism="Aspergillus niger"
                     /mol_type="mRNA"
                     /db_xref="taxon:5061"
                     /clone="3235"
                     /lab_host="E. coli"
                     /clone_lib="Aspergillus niger, pYES2 (XhoI-EcoRI)"
                     /note="Vector: pYES2; Site_1: Xho-I; Site_2: EcoRI; cDNA
                     was synthesized with ZAP kit (Stratagene) using poly(A)+
                     RNA isolated from Aspergillus niger grown in both complete
                     and minimal media. Synthesis was primed with oligo(dT)
                     primer/XhoI-linker. EcoRI adaptors were later ligated to
                     polished ends. EcoRI-XhoI-digested cDNA was ligated with
                     EcoRI-XhoI-digested pYES2 (Invitrogen Corp). This vector
                     permits expression of our library in yeast. "
ORIGIN

Alignment Scores:
Pred. No.:              2.51e+03      Length:         438
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     2             Gaps:           0

US-09-858-332G-28 (1-6) x BE758972 (1-438)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        130 CACAACCACAACCACAAC 147

RESULT 353
CF986640
LOCUS       CF986640                438 bp    mRNA    linear   EST 04-FEB-2005
DEFINITION  23319rsicee_2448.y1 Oryza sativa cv. LYP9 tillering whole plant
            cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF986640
VERSION     CF986640.1  GI:58588332
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 438)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
  PUBMED    15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 438
            POLYA=No.
FEATURES             Location/Qualifiers
```

```
     source          1. .438
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="tillering"
                     /clone_lib="Oryza sativa cv. LYP9 tillering whole plant
                     cDNA library"
ORIGIN

Alignment Scores:
Pred. No.:              2.51e+03        Length:         438
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     7               Gaps:           0

US-09-858-332G-28 (1-6) x CF986640 (1-438)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        219 CACAACCACAACCACAAC 236

RESULT 354
CK080515
LOCUS       CK080515                438 bp    mRNA    linear   EST 05-FEB-2005
DEFINITION  81076rsicen_7204.y1 Oryza sativa cv. LYP9 booting whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CK080515
VERSION     CK080515.1  GI:58691828
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 438)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 438
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .438
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="booting"
                     /clone_lib="Oryza sativa cv. LYP9 booting whole plant cDNA
                     library"
ORIGIN

Alignment Scores:
Pred. No.:              2.51e+03        Length:         438
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     7               Gaps:           0

US-09-858-332G-28 (1-6) x CK080515 (1-438)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        204 CACAACCACAACCACAAC 221

RESULT 355
CZ590712/c
LOCUS       CZ590712                438 bp    DNA     linear   GSS 20-JUN-2005
DEFINITION  OA_BBa0185I02.r OA_BBa Oryza alta genomic clone OA_BBa0185I02 3',
            genomic survey sequence.
ACCESSION   CZ590712
VERSION     CZ590712.1  GI:68048051
KEYWORDS    GSS.
SOURCE      Oryza alta
  ORGANISM  Oryza alta
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 438)
  AUTHORS   SanMiguel,P., Westerman,R., Kim,H., Yu,Y., Wissotski,M., Yost,D.,
            Stum,D., Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C.,
            Hatfield,J., Soderlund,C., Wing,R. and Jackson,S.A.
  TITLE     OMAP (Oryza Map Alignment Project)- Purdue University
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Scott A. Jackson
            Jackson Laboratory
            Purdue University
            915 W. State St., West Lafayette, IN 47907, USA
            Tel: 7654963621
            Fax: 7654967255
            Email: sjackson@purdue.edu
            Basecalling by phred version 0.020425.c. This sequence was derived
            from the raw sequence read by clipping with lucy version 1.19s.
            Bases 99-536 of the raw sequence (length 1102) were retained after
            clipping.
            Plate: 0185  row: I  column: 02
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .438
                     /organism="Oryza alta"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:52545"
                     /clone="OA_BBa0185I02"
                     /tissue_type="young leaves"
                     /lab_host="DH10B-T1 phage resistant"
                     /clone_lib="OA_BBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

Alignment Scores:
Pred. No.:              2.51e+03        Length:         438
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     10              Gaps:           0

US-09-858-332G-28 (1-6) x CZ590712 (1-438)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        382 CACAACCACAACCACAAC 365

RESULT 356
```

```
AW596891
LOCUS       AW596891                440 bp    mRNA    linear   EST 14-JUL-2004
DEFINITION  sj84b04.y1 Gm-c1034 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1034-1640 5' similar to SW:IF4V_TOBAC Q40465 EUKARYOTIC
            INITIATION FACTOR 4A-11 ;, mRNA sequence.
ACCESSION   AW596891
VERSION     AW596891.1  GI:7284297
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 440)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,
            Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
            Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
            Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
            Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            When it has been determined, an EST from the other end of this
            clone is listed in the 'Other ESTs on clone' field. This clone is
            available through: Biogenetic Services, 801 32nd Ave. Brookings, SD
            57006 USA (phone: 800 423 4163; email: info@biogeneticservices.com)
            Insert Length: 1578   Std Error: 0.00.
FEATURES             Location/Qualifiers
     source          1. .440
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /cultivar="Williams"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1034-1640"
                     /tissue_type="very young seed coats of greenhouse grown
                     plants"
                     /lab_host="DH10B"
                     /clone_lib="Gm-c1034"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; This
                     cDNA library was constructed from mRNA isolated from very
                     young seed coats (20-50mgs fresh weight) of greenhouse
                     grown plants. The library was prepared using the Life
                     Technologies pSuperScript cDNA library construction kit.
                     Complementary DNA was synthesized from mRNA using a
                     poly(dT) sequence with a NotI restrictions site. SalI
                     linkers adapters were ligated to the blunt-ended cDNA
                     fragments followed by NotI digestion. The cDNA fragments
                     were directionally cloned into the NotI-SalI restriction
                     site of the pSPORT1 vector. The ligated cDNA fragments
                     were  transformed into E.coli ElectroMax DH10B host cells.
                     This library  was constructed by Dr. Lila Vodkin and Dr.
                     Anu Khanna."
ORIGIN

Alignment Scores:
Pred. No.:                2.53e+03        Length:          440
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       1               Gaps:            0

US-09-858-332G-28 (1-6) x AW596891 (1-440)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         40 CACAACCACAACCACAAC 57

RESULT 357
BE461297
LOCUS       BE461297                440 bp    mRNA    linear   EST 18-MAY-2001
DEFINITION  EST412716 tomato breaker fruit, TIGR Lycopersicon esculentum cDNA
            clone cLEG38K17, mRNA sequence.
ACCESSION   BE461297
VERSION     BE461297.1  GI:9505599
KEYWORDS    EST.
SOURCE      Lycopersicon esculentum (Solanum lycopersicum)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 440)
  AUTHORS   Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
            Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
            Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and
            Tanksley,S.D.
  TITLE     Generation of ESTs from tomato fruit tissue, breaker stage
  JOURNAL   Unpublished (2000)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html
            5 prime sequence.
FEATURES             Location/Qualifiers
     source          1. .440
                     /organism="Lycopersicon esculentum"
                     /mol_type="mRNA"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cLEG38K17"
                     /tissue_type="Pericarp"
                     /dev_stage="breaker"
                     /lab_host="SOLR"
                     /clone_lib="tomato breaker fruit, TIGR"
                     /note="Vector: pBluescriptSKmCUadapt; Site_1: EcoR1;
                     Site_2: XhoI; Fruit were harvested at the breaker stage
                     (first sign of lycopene accumulation on the blossom end of
                     the fruit).  Fruit were cut in half and the seeds and
                     locules were discarded prior to freezing the pericarp."
ORIGIN

Alignment Scores:
Pred. No.:                2.53e+03        Length:          440
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       2               Gaps:            0

US-09-858-332G-28 (1-6) x BE461297 (1-440)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        421 CACAATCACAATCACAAT 438

RESULT 358
BQ804984
LOCUS       BQ804984                440 bp    mRNA    linear   EST 31-JUL-2002
DEFINITION  WHE3561_D04_H07ZS Wheat developing grains cDNA library Triticum
            aestivum cDNA clone WHE3561_D04_H07, mRNA sequence.
ACCESSION   BQ804984
VERSION     BQ804984.1  GI:22029193
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
```

Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 440)
AUTHORS Altenbach,S., Anderson,O.D., Chao,S., Chin,A., Close,T.J., Cronin,K., Crossman,C., Fenton,R.D., Lazo,G.R., Pham,J., Rausch,C.J., Wilson,C. and Woo,J.
TITLE The structure and function of the expressed portion of the wheat genomes - Developing grains cDNA library
JOURNAL Unpublished (2002)
COMMENT Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oandersn@pw.usda.gov
Sequences have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
Seq primer: SK primer.

FEATURES Location/Qualifiers
source 1. .440
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Butte 86"
/db_xref="taxon:4565"
/clone="WHE3561_D04_H07"
/tissue_type="whole grains"
/dev_stage="3-44 days post anthesis seed"
/lab_host="E. coli SOLR"
/clone_lib="Wheat developing grains cDNA library"
/note="Vector: Lambda ZAP II, excised phagemid; Site_1: EcoRI; Plants were grown under six following different environmental regimes in greenhouse, Environment 1) 24oC/17oC day/night, well-watered, with post-anthesis fertilizer, Environment 2) 24oC/17oC day/night, well-watered, without post-anthesis fertilizer, Environment 3) 37oC/17oC day/night, well-watered, with post-anthesis fertilizer, Environment 4) 37oC/17oC day/night, well-watered, without post-anthesis fertilizer, Environment 5) 37oC/17oC day/night plus drought, with post-anthesis fertilizer, Environment 6) 37oC/17oC day/night plus drought, without post-anthesis fertilizer, developing wheat grains from the following were excised and frozen in liquid nitrogen, Environment 1 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44 DPA Environment 2 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44 DPA Environment 3 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 34 DPA Environment 4 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 34 DPA Environment 5 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 30 DPA Environment 6 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 30 DPA and total RNA was prepared by S. Altenbach and K. Cronin at USDA-ARS, Albany, CA. A cDNA library was made using poly (A) RNA, and the cDNA clones were in vivo excised to give pBluescript SK(-) phagemids in the TJ Close lab (Chin, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (others)."

ORIGIN

Alignment Scores:

Pred. No.:	2.53e+03	Length:	440
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	5	Gaps:	0

US-09-858-332G-28 (1-6) x BQ804984 (1-440)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        334 CACAACCACAACCACAAT 351
```

RESULT 359
H18998
LOCUS H18998 441 bp mRNA linear EST 29-JUN-1995
DEFINITION yn51h01.s1 Soares adult brain N2b5HB55Y Homo sapiens cDNA clone IMAGE:171985 3', mRNA sequence.
ACCESSION H18998
VERSION H18998.1 GI:885238
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 441)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1060
High quality sequence stops: 294 Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1060 Std Error: 0.00
Seq primer: Promega -21m13
High quality sequence stop: 294.

FEATURES Location/Qualifiers
source 1. .441
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:3834020"
/db_xref="taxon:9606"
/clone="IMAGE:171985"
/sex="Male"
/dev_stage="55-year old"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares adult brain N2b5HB55Y"
/note="Organ: brain; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGCTTTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 53. Library constructed by Bento Soares and M.Fatima Bonaldo. The adult brain RNA was provided by Dr. Donald H. Gilden. Tissue was acquired 17-18 hours after death which occurred in consequence of a ruptured aortic aneurysm. RNA was prepared from a pool of tissues representing the following areas of the brain: frontal, parietal, temporal and occipital cortex from the left and right hemispheres, subcortical white matter, basal ganglia, thalamus, cerebellum, midbrain, pons and medulla."

ORIGIN

Alignment Scores:

Pred. No.:	2.53e+03	Length:	441
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	8	Gaps:	0

US-09-858-332G-28 (1-6) x H18998 (1-441)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        191 CATAATCATAATCATAAT 208
```

RESULT 360
BM577130
LOCUS BM577130 442 bp mRNA linear EST 22-FEB-2002
DEFINITION 17000687163345 A.Gam.ad.cDNA.blood1 Anopheles gambiae cDNA clone 19600449713077 5', mRNA sequence.
ACCESSION BM577130
VERSION BM577130.1 GI:18865597
KEYWORDS EST.
SOURCE Anopheles gambiae (African malaria mosquito)
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae; Anophelinae; Anopheles.
REFERENCE 1 (bases 1 to 442)
AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
TITLE Celera Anopheles gambiae EST project
JOURNAL Unpublished (2002)
COMMENT Contact: Holt R.A.
Celera Genomics
45 w. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltRA@celera.com
Plate: NU01004AWU row: J column: 07
Seq primer: M13 Reverse.
FEATURES Location/Qualifiers
source 1. .442
/organism="Anopheles gambiae"
/mol_type="mRNA"
/strain="RSP-ST (Reduced susc. to Permethrin - std. chromosome)"
/db_xref="taxon:7165"
/clone="19600449713077"
/dev_stage="Adult"
/lab_host="DH10b"
/clone_lib="A.Gam.ad.cDNA.blood1"
/note="Vector: pSport1; Site_1: Sal1; Site_2: Not1; Whole adult mosquitoes (mixed sex) frozen on liquid nitrogen 24 hours after human blood feeding. cDNA inserts >500 bp cloned directionally into pSport 1. Not 1 site is 3'. Clones available through the Malaria Research and Reference Reagent Resource Center (www.malaria.mr4.org)"
ORIGIN

Alignment Scores:

Pred. No.:	2.54e+03	Length:	442
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	3	Gaps:	0

US-09-858-332G-28 (1-6) x BM577130 (1-442)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        218 CACAACCACAACCACAAC 235
```

RESULT 361
CF988556
LOCUS CF988556 442 bp mRNA linear EST 04-FEB-2005
DEFINITION 18328rsicee_9466.y1 Oryza sativa cv. LYP9 tillering whole plant cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA sequence.
ACCESSION CF988556
VERSION CF988556.1 GI:58590248
KEYWORDS EST.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 442)
AUTHORS Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W., Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X., Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J., Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q., Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z., Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P., Wu,S. and Liu,J.
TITLE The Genomes of Oryza sativa: A History of Duplications
JOURNAL PLoS Biol. 3 (2), e38 (2005)
PUBMED 15685292
COMMENT Contact: Yan Zhou
Bioinformatics Department
Hangzhou Genomics Institute
No.51 Zhijiang Road, Hangzhou 310008, China
Tel: 86-571-56805886
Fax: 86-571-56805884
Email: zhouyan@genomics.org.cn
Seq primer: M13 Forward
High quality sequence stop: 442
POLYA=No.
FEATURES Location/Qualifiers
source 1. .442
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/cultivar="LYP9"
/db_xref="taxon:39946"
/tissue_type="whole plant"
/dev_stage="tillering"
/clone_lib="Oryza sativa cv. LYP9 tillering whole plant cDNA library"
ORIGIN

Alignment Scores:

Pred. No.:	2.54e+03	Length:	442
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	7	Gaps:	0

US-09-858-332G-28 (1-6) x CF988556 (1-442)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        199 CACAACCACAACCACAAC 216
```

RESULT 362
CN193991/c
LOCUS CN193991 442 bp mRNA linear EST 05-APR-2004
DEFINITION rg99g11.y1 Meloidogyne hapla female SMART pGEM Meloidogyne hapla cDNA 5', mRNA sequence.
ACCESSION CN193991
VERSION CN193991.1 GI:46218930
KEYWORDS EST.
SOURCE Meloidogyne hapla
ORGANISM Meloidogyne hapla
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina; Tylenchoidea; Meloidogynidae; Meloidogyninae; Meloidogyne.
REFERENCE 1 (bases 1 to 442)
AUTHORS McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagareishvili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe,M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.

```
  TITLE     The Washington Univ. Nematode EST Project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: McCarter JP
            The Washington Univ. Nematode EST Project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Cloned unidirectionally. Poly(A)+ RNA was concentrated and purified
            using Dynabeads (Dynal) and mRNA eluted for first strand synthesis.
            First strand cDNA was created using MMLV RT (Powerscript, Clontech)
            and primed with oligo(dT) with XhoI site and 5'SMART 'anchor' added
            using chimeric DNA-RNA oligo. 12 PCR cycles were done using first
            strand and primers specific to SMART oligo and 3' end. Double
            stranded cDNA was digested using XhoI/NotI, fractioned on
            Chroma-spin 400 columns (Clontech) and ligated to digested
            pGEM-11zf(+) plasmid. Chemically competent DH10B cells were used as
            host cells. Library materials provided by Dr. David Bird of North
            Carolina State University. Library construction by Jeff Rousch. See
            www.nematode.net for additional project information.
            Putative full length read
            The vector to vector length is 443
            Seq primer: Sp6.
FEATURES                Location/Qualifiers
     source             1. .442
                        /organism="Meloidogyne hapla"
                        /mol_type="mRNA"
                        /db_xref="taxon:6305"
                        /tissue_type="whole organism"
                        /dev_stage="adult"
                        /lab_host="DH10B"
                        /clone_lib="Meloidogyne hapla female SMART pGEM"
                        /note="Vector: plasmid (ampicillin resistant); Site_1:
                        XhoI; Site_2: NotI; Cloned unidirectionally. Poly(A)+ RNA
                        was concentrated and purified using Dynabeads (Dynal) and
                        mRNA eluted for first strand synthesis. First strand cDNA
                        was created using MMLV RT (Powerscript, Clontech) and
                        primed with oligo(dT) with XhoI site and 5'SMART 'anchor'
                        added using chimeric DNA-RNA oligo. 12 PCR cycles were
                        done using first strand and primers specific to SMART
                        oligo and 3' end. Double stranded cDNA was digested using
                        XhoI/NotI, fractioned on Chroma-spin 400 columns
                        (Clontech) and ligated to digested pGEM-11zf(+) plasmid.
                        Chemically competent DH10B cells were used as host cells.
                        Library materials provided by Dr. David Bird of North
                        Carolina State University. Library construction by Jeff
                        Rousch. See www.nematode.net for additional project
                        information."
ORIGIN

Alignment Scores:
Pred. No.:                2.54e+03        Length:         442
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       7               Gaps:           0

US-09-858-332G-28 (1-6) x CN193991 (1-442)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        278 CATAATCACAATCATAAT 261

RESULT 363
BE604997
LOCUS       BE604997                  444 bp    mRNA    linear   EST 21-AUG-2000
DEFINITION  WHE1713-1716_E09_E09ZS Wheat heat stressed spike cDNA library
            Triticum aestivum cDNA clone WHE1713-1716_E09_E09, mRNA sequence.
ACCESSION   BE604997
VERSION     BE604997.1  GI:9862267
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 444)
  AUTHORS   Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D.,
            Han,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Malatrasi,M., Miller,R.,
            Nguyen,H.T., Rausch,C.J., Seaton,C.L., Tong,J.C. and Zhang,D.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Heat stressed spike cDNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene SK primer.
FEATURES                Location/Qualifiers
     source             1. .444
                        /organism="Triticum aestivum"
                        /mol_type="mRNA"
                        /cultivar="Chinese Spring"
                        /db_xref="taxon:4565"
                        /clone="WHE1713-1716_E09_E09"
                        /tissue_type="Whole spike"
                        /dev_stage="Spikes at 5, 10, 15 and 20 days after
                        anthesis"
                        /lab_host="E. coli SOLR"
                        /clone_lib="Wheat heat stressed spike cDNA library"
                        /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                        Site_1: EcoRI; Site_2: XhoI; Spikes at 5, 10, 15 and 20
                        days after anthesis were heat stressed under  two
                        conditions at Texas Tech University (D. Zhang in HT Nguyrn
                        lab):  (1) at 38 C for 4 hours and (2) 5 days of cyclic
                        treatment of 38 C  for 4 hours.  Total RNA and poly(A) RNA
                        were prepared, a cDNA library  was made, and the cDNA
                        clones were in vivo excised to give pBluescript  phagemids
                        in the TJ Close lab (Choi, Close, Fenton, Malatrasi) at
                        the  University of California, Riverside.  Plasmid DNA
                        preparations and  DNA sequencing were performed in the OD
                        Anderson lab (all other authors)."
ORIGIN

Alignment Scores:
Pred. No.:                2.55e+03        Length:         444
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       2               Gaps:           0

US-09-858-332G-28 (1-6) x BE604997 (1-444)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        337 CACAACCACAACCACAAT 354

RESULT 364
AA391089
LOCUS       AA391089                  445 bp    mRNA    linear   EST 19-APR-2001
DEFINITION  LD10018.5prime LD Drosophila melanogaster embryo BlueScript
            Drosophila melanogaster cDNA clone LD10018 5prime, mRNA sequence.
ACCESSION   AA391089
VERSION     AA391089.1  GI:2044066
KEYWORDS    EST.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
```

```
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 445)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Blast expect value = 2.6E-69 on X62639: D.melanogaster mRNA for
            hrp48.1
            Plate: 100  row: B  column: 6
            High quality sequence stop: 319.
FEATURES             Location/Qualifiers
     source          1. .445
                     /organism="Drosophila melanogaster"
                     /mol_type="mRNA"
                     /db_xref="BDGP_EST:BDcln009293"
                     /db_xref="taxon:7227"
                     /clone="LD10018"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="SOLR"
                     /clone_lib="LD Drosophila melanogaster embryo BlueScript"
                     /note="Organ: embryo; Vector: BlueScript SK; Site_1:
                     EcoRI; Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
                     Synthesis kit. Oligo dT-primed and directionally cloned at
                     EcoRI and XhoI in BlueScript SK(+/-)"
ORIGIN

Alignment Scores:
Pred. No.:              2.56e+03        Length:         445
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AA391089 (1-445)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        309 CACAACCACAACCACAAC 326

RESULT 365
AL379537/c
LOCUS       AL379537                446 bp    mRNA    linear   EST 03-AUG-2000
DEFINITION  MtBB45H08R1 MtBB Medicago truncatula cDNA clone MtBB45H08 T7, mRNA
            sequence.
ACCESSION   AL379537
VERSION     AL379537.1  GI:9679289
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 446)
  AUTHORS   Journet,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jaillon,O.,
            Niebel,A., Carreau,V., Chatagnier,O., Kahn,D.,
            Gianinazzi-Pearson,V. and Gamas,P.
  TITLE     Medicago truncatula ESTs from Sinorhizobium meliloti-induced root
            nodules
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de
            Biologie Moleculaire des Relations Plantes-Microorganismes,
            CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :
            Mt-est@toulouse.inra.fr Website :
            http://sequence.toulouse.inra.fr/Mtruncatula.html).
FEATURES             Location/Qualifiers
     source          1. .446
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /cultivar="Jemalong"
                     /db_xref="taxon:3880"
                     /clone="MtBB45H08"
                     /tissue_type="symbiotic root nodules"
                     /dev_stage="harvested 4 days post inoculation with
                     Sinorhizobium meliloti"
                     /clone_lib="MtBB"
                     /note="Vector: pBluescript pSK; Site_1: EcoRI; Site_2:
                     XhoI; Plants were grown in an aeroponic chamber on
                     nitrogen-rich medium for 21 days. Three days before
                     inoculation with Sinorhizobium meliloti, the medium was
                     replaced by N-free medium. Root nodules (+ short adjacent
                     root segments) were harvested 4 days post inoculation.
                     cDNA was prepared from polyA+ enriched RNA. The cDNA was
                     directionally ligated into Uni-zap XR vector from
                     Stratagene and packaged using Gigapack Gold packaging
                     extracts. Plasmids containing cDNA inserts were
                     mass-excised from phage stocks using ExAssit helper phage
                     and propagated in SOLR cells. Clone ordering and
                     sequencing was performed by the Centre National de
                     Sequencage (Genoscope, Evry, France)."
ORIGIN

Alignment Scores:
Pred. No.:              2.56e+03        Length:         446
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AL379537 (1-446)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        259 CATAATCATAATCATAAT 242

RESULT 366
BJ324275
LOCUS       BJ324275                446 bp    mRNA    linear   EST 05-MAR-2002
DEFINITION  BJ324275 Dictyostelium discoideum cDNA library, AF Dictyostelium
            discoideum cDNA clone dda5f17 5', mRNA sequence.
ACCESSION   BJ324275
VERSION     BJ324275.1  GI:19154405
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 446)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the aggregation
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .446
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
```

/strain="AX4"
/db_xref="taxon:44689"
/clone="dda5f17"
/sex="mat A"
/dev_stage="Aggregation stage"
/clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN

Alignment Scores:

Pred. No.:	2.56e+03	Length:	446
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	3	Gaps:	0

US-09-858-332G-28 (1-6) x BJ324275 (1-446)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        233 CATAACCATAACCATAAT 250
```

RESULT 367
BJ434889/c
LOCUS BJ434889 446 bp mRNA linear EST 13-MAR-2002
DEFINITION BJ434889 Dictyostelium discoideum cDNA library, VF Dictyostelium discoideum cDNA clone ddv25f07 3', mRNA sequence.
ACCESSION BJ434889
VERSION BJ434889.1 GI:19409611
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 446)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the vegetative stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES Location/Qualifiers
source 1. .446
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddv25f07"
/sex="mat A"
/dev_stage="Growth phase"
/clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:

Pred. No.:	2.56e+03	Length:	446
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	3	Gaps:	0

US-09-858-332G-28 (1-6) x BJ434889 (1-446)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        149 CACAATCACAATCACAAT 132
```

RESULT 368
CA613496/c
LOCUS CA613496 446 bp mRNA linear EST 23-NOV-2002
DEFINITION wr1.pk0146.h2 wr1 Triticum aestivum cDNA clone wr1.pk0146.h2 5' end, mRNA sequence.
ACCESSION CA613496
VERSION CA613496.1 GI:25191793
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 446)
AUTHORS Tingey,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z., Miao,G., Caraher,N. and Hanafey,M.K.
TITLE DuPont Wheat cDNA Sequence
JOURNAL Unpublished (2002)
COMMENT Contact: Scott V. Tingey
Crop Genetics
E. I. DuPont de Nemours and Company
1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
Tel: 302-631-2602
Fax: 302-631-2607
Email: Scott.V.Tingey@USA.dupont.com
Seq primer: M13.
FEATURES Location/Qualifiers
source 1. .446
/organism="Triticum aestivum"
/mol_type="mRNA"
/db_xref="taxon:4565"
/clone="wr1.pk0146.h2"
/tissue_type="root"
/clone_lib="wr1"
/note="Vector: pBluescript SK+; Site_1: EcoRI; Site_2: XhoI; Wheat (Triticum aestivum L.) root; 7 day old seedling, light grown"
ORIGIN

Alignment Scores:

Pred. No.:	2.56e+03	Length:	446
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-09-858-332G-28 (1-6) x CA613496 (1-446)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        242 CATAATCATAATCATAAC 225
```

RESULT 369
AA636187
LOCUS AA636187 447 bp mRNA linear EST 22-OCT-1997
DEFINITION vn15e07.r1 Stratagene mouse heart (#937316) Mus musculus cDNA clone IMAGE:1021284 5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION AA636187
VERSION AA636187.1 GI:2559966
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 447)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)

```
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:572060
            Seq primer: -28m13 rev1 ET from Amersham.
FEATURES              Location/Qualifiers
     source           1. .447
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /strain="NIH Swiss"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:1021284"
                      /sex="pooled"
                      /tissue_type="heart"
                      /dev_stage="13 day embryos"
                      /lab_host="SOLR (kanamycin resistant)"
                      /clone_lib="Stratagene mouse heart (#937316)"
                      /note="Organ: heart; Vector: pBluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                      Oligo dT. 93 pooled NIH/Swiss 13 day embryo hearts.
                      Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5'
                      adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor
                      sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
ORIGIN

Alignment Scores:
Pred. No.:              2.57e+03        Length:         447
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AA636187 (1-447)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        244 CACAACCATAACCACAAC 261

RESULT 370
CA410761
LOCUS       CA410761                447 bp    mRNA    linear   EST 01-JAN-2003
DEFINITION  E031_F -P normal and proteoid roots 7 and 10 DAE Lupinus albus cDNA
            clone E031 5', mRNA sequence.
ACCESSION   CA410761
VERSION     CA410761.1  GI:27459765
KEYWORDS    EST.
SOURCE      Lupinus albus (white lupine)
  ORGANISM  Lupinus albus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Genisteae;
            Lupinus.
REFERENCE   1  (bases 1 to 447)
  AUTHORS   Uhde-Stone,C., Zinn,K.E., Ramirez-Yanez,M., Li,A., Vance,C.P. and
            Allan,D.L.
  TITLE     Nylon filter arrays reveal differential gene expression in proteoid
            roots of white lupin in response to P deficiency
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Uhde-Stone, Claudia
            Department of Agronomy and Plant Genetics
            University of Minnesota
            439 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA
            Tel: 612 624 6765
            Fax: 612 625 2208
            Email: cuhde@soils.umn.edu
            PCR PRimers
            FORWARD: T3
            BACKWARD: T7
            Seq primer: AATTAACCCTCACTAAAGGG.
FEATURES              Location/Qualifiers
     source           1. .447
                      /organism="Lupinus albus"
                      /mol_type="mRNA"
                      /cultivar="Ultra"
                      /db_xref="taxon:3870"
                      /clone="E031"
                      /dev_stage="7 and 10 DAE"
                      /clone_lib="-P normal and proteoid roots 7 and 10 DAE"
                      /note="Organ: -P normal and proteoid root (cluster root);
                      Vector: pBLUESCRIPT SK(+/-); Site_1: EcoRI; Site_2: XhoI;
                      Normal and proteoid roots of white lupin grown under P
                      deficient conditions were harvested 7 and 10 DAE (days
                      after emergence), and poly(A)+ RNA was isolated. The
                      poly(A)+ RNA obtained from plants 7 and 10 DAE was
                      combined in a 1:1 ratio and 7 ug total RNA was used for
                      reconstruction of a proteoid root cDNA library in the
                      phage ZAPII vector according to the manufacturer's
                      instructions (Stratagene, La Jolla, CA, USA). For
                      conversion of the phage library into the plasmid form,
                      mass excision was performed according to the procedure
                      described by Stratagene."
ORIGIN

Alignment Scores:
Pred. No.:              2.57e+03        Length:         447
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x CA410761 (1-447)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        238 CACAACCACAACCACAAC 255

RESULT 371
BE580088
LOCUS       BE580088                448 bp    mRNA    linear   EST 09-MAY-2001
DEFINITION  kq37a01.y1 TBN95TM-SSR Strongyloides stercoralis cDNA 5' similar to
            contains element MER11 repetitive element ;, mRNA sequence.
ACCESSION   BE580088
VERSION     BE580088.1  GI:9831030
KEYWORDS    EST.
SOURCE      Strongyloides stercoralis
  ORGANISM  Strongyloides stercoralis
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Panagrolaimoidea; Strongyloididae; Strongyloides.
REFERENCE   1  (bases 1 to 448)
  AUTHORS   McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J.,
            Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B.,
            Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C.,
            Tsagareishvili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C.,
            Underwood,K., Steptoe,M., Allen,M., Person,B., Swaller,T.,
            Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     The Washington Univ. Nematode EST Project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: McCarter JP
            The Washington Univ. Nematode EST Project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            The library was constructed by Dr. Thomas Nutman and colleagues of
            NIAID, NIH (tnutman@nih.gov). DNA Sequencing by: Washington
            University Genome Sequencing Center St. Louis.
```

```
            High quality sequence stop: 445.
FEATURES             Location/Qualifiers
     source          1. .448
                     /organism="Strongyloides stercoralis"
                     /mol_type="mRNA"
                     /strain="Rhabditiform larvae obtained from gerbils"
                     /db_xref="taxon:6248"
                     /lab_host="XL-1 Blue MRF' (Stratagene)"
                     /clone_lib="TBN95TM-SSR"
                     /note="Vector: Lambda Uni-ZAP XR (Stratagene); Site_1:
                     EcoRI; Site_2: XhoI; mRNA was purified from 2 x 10E3
                     rhabditiform larvae which had been isolated from gerbils
                     experimentally infected with larvae originally isolated
                     from experimentally infected dogs. cDNA was constructed
                     and, using adaptors, was cloned unidirectionally into the
                     vector from the EcoRI site to the XhoI site. The library
                     has an unamplified titer of 1 x 10E5 pfu/ml and an
                     amplified, undiluted titer of 9 x 10E11 pfu/ml. The
                     average insert size of the unamplifed library is 675 bp
                     (range, 100-1700)."
ORIGIN

Alignment Scores:
Pred. No.:                2.57e+03        Length:         448
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       2               Gaps:           0

US-09-858-332G-28 (1-6) x BE580088 (1-448)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        164 CACAATCACAACCACAAC 181

RESULT 372
CF965703
LOCUS       CF965703                448 bp    mRNA    linear   EST 15-JAN-2004
DEFINITION  14418rsiced_19030.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF965703
VERSION     CF965703.1  GI:38481614
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 448)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 448
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .448
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="trefoil"
                     /clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA
                     library"
ORIGIN

Alignment Scores:
Pred. No.:                2.57e+03        Length:         448
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       7               Gaps:           0

US-09-858-332G-28 (1-6) x CF965703 (1-448)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        130 CACAACCACAACCACAAC 147

RESULT 373
AQ778216/c
LOCUS       AQ778216                448 bp    DNA     linear   GSS 02-AUG-1999
DEFINITION  HS_3206_A1_F06_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3206 Col=11 Row=K, genomic survey
            sequence.
ACCESSION   AQ778216
VERSION     AQ778216.1  GI:5681176
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 448)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 3206  row: K  column: 11
            Seq primer: M13 Reverse
            Class: BAC ends
            High quality sequence stop: 448.
FEATURES             Location/Qualifiers
     source          1. .448
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=3206 Col=11 Row=K"
                     /sex="male"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
ORIGIN

Alignment Scores:
Pred. No.:                2.57e+03        Length:         448
```

```
Score:                 42.00        Matches:       6
Percent Similarity:    100.00%      Conservative:  0
Best Local Similarity: 100.00%      Mismatches:    0
Query Match:           100.00%      Indels:        0
DB:                    9            Gaps:          0

US-09-858-332G-28 (1-6) x AQ778216 (1-448)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         87 CATAATCATAATCATAAT 70

RESULT 374
AI317046
LOCUS       AI317046                449 bp    mRNA    linear   EST 17-DEC-1998
DEFINITION  uk71g09.y1 Schiller mouse AtT20 Mus musculus cDNA clone
            IMAGE:1974496 5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0
            KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION   AI317046
VERSION     AI317046.1  GI:4032313
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:991236
            Putative full length read
            vector to vector length is 593
            Seq primer: -40RP from Gibco
            High quality sequence stop: 399.
FEATURES             Location/Qualifiers
     source          1. .449
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1974496"
                     /cell_line="pituitary cell line"
                     /lab_host="SOLR"
                     /clone_lib="Schiller mouse AtT20"
                     /note="Organ: pituitary; Vector: pBluescript SK-
                     (Stratagene); Site_1: EcoRI; Site_2: XhoI; Double-stranded
                     cDNA was prepared from cell line AtT-20 using primer
                     5'-GAGAGAGAGAGAGAGAGAAACTAGTCTGAGT(18)-3'.  An EcoRI
                     adaptor was used on the 5' end of the cDNA as follows:
                     5'-AATTCGGCACGAG-3'.  The library was size-selected and
                     went through one round of amplification.  Average insert
                     size is 1.7 kb, with a range from 0.4-12 kb.  This library
                     was constructed by Dr. Martin Schiller (Johns Hopkins
                     University). "
ORIGIN

Alignment Scores:
Pred. No.:             2.58e+03     Length:        449
Score:                 42.00        Matches:       6
Percent Similarity:    100.00%      Conservative:  0
Best Local Similarity: 100.00%      Mismatches:    0
Query Match:           100.00%      Indels:        0
DB:                    1            Gaps:          0

US-09-858-332G-28 (1-6) x AI317046 (1-449)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        249 CACAACCATAACCACAAC 266

RESULT 375
BY059656
LOCUS       BY059656                449 bp    mRNA    linear   EST 06-DEC-2002
DEFINITION  BY059656 RIKEN full-length enriched, 16 days embryo heart Mus
            musculus cDNA clone I920002M19 5', mRNA sequence.
ACCESSION   BY059656
VERSION     BY059656.1  GI:26165104
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
            Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
            Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
            Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
            Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
            Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
            Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
            Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
            Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
            Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
            Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
            Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
            Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
            Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
            Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
            Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
            Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
            Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
            Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
            Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
            Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
            Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
            Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
            Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
            Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
            Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
   PUBMED   12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
            Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
            Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,
            Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,
            Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and
            Hayashizaki,Y. Direct Submission
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
```

genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

FEATURES Location/Qualifiers
source 1. .449
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="I920002M19"
/tissue_type="heart"
/dev_stage="16 days embryo"
/clone_lib="RIKEN full-length enriched, 16 days embryo heart"

ORIGIN

Alignment Scores:

Pred. No.:	2.58e+03	Length:	449
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	5	Gaps:	0

US-09-858-332G-28 (1-6) x BY059656 (1-449)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        295 CACAACCATAACCACAAC 312
```

RESULT 376
CF960519
LOCUS CF960519 449 bp mRNA linear EST 15-JAN-2004
DEFINITION 12399rsiced_7134.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA sequence.
ACCESSION CF960519
VERSION CF960519.1 GI:38476386
KEYWORDS EST.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 449)
AUTHORS Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W., Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X., Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J., Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q., Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z., Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P., Wu,S. and Liu,J.
TITLE The Genomes of Oryza sativa: A History of Duplications
JOURNAL PLoS Biol. 3 (2), e38 (2005)
PUBMED 15685292
COMMENT Contact: Yan Zhou
Bioinformatics Department
Hangzhou Genomics Institute
No.51 Zhijiang Road, Hangzhou 310008, China
Tel: 86-571-56805886
Fax: 86-571-56805884
Email: zhouyan@genomics.org.cn
Seq primer: M13 Forward
High quality sequence stop: 449
POLYA=No.

FEATURES Location/Qualifiers
source 1. .449
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/cultivar="LYP9"
/db_xref="taxon:39946"
/tissue_type="whole plant"
/dev_stage="trefoil"
/clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA library"

ORIGIN

Alignment Scores:

Pred. No.:	2.58e+03	Length:	449
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	7	Gaps:	0

US-09-858-332G-28 (1-6) x CF960519 (1-449)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        200 CACAACCACAACCACAAC 217
```

RESULT 377
CF963362
LOCUS CF963362 450 bp mRNA linear EST 15-JAN-2004
DEFINITION 13642rsiced_5972.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA sequence.
ACCESSION CF963362
VERSION CF963362.1 GI:38479229
KEYWORDS EST.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 450)
AUTHORS Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W., Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X., Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J., Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q., Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z., Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P., Wu,S. and Liu,J.
TITLE The Genomes of Oryza sativa: A History of Duplications
JOURNAL PLoS Biol. 3 (2), e38 (2005)
PUBMED 15685292
COMMENT Contact: Yan Zhou
Bioinformatics Department
Hangzhou Genomics Institute
No.51 Zhijiang Road, Hangzhou 310008, China
Tel: 86-571-56805886
Fax: 86-571-56805884
Email: zhouyan@genomics.org.cn
Seq primer: M13 Forward
High quality sequence stop: 450
POLYA=No.

FEATURES Location/Qualifiers
source 1. .450
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/cultivar="LYP9"
/db_xref="taxon:39946"
/tissue_type="whole plant"
/dev_stage="trefoil"
/clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA

library"
ORIGIN

Alignment Scores:
Pred. No.: 2.59e+03 Length: 450
Score: 42.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0

US-09-858-332G-28 (1-6) x CF963362 (1-450)

Qy 1 HisAsnHisAsnHisAsn 6
||||||||||||||||||
Db 197 CACAACCACAACCACAAC 214

RESULT 378
BQ244558
LOCUS BQ244558 451 bp mRNA linear EST 03-MAY-2002
DEFINITION TaE15040H11R TaE15 Triticum aestivum cDNA clone TaE15040H11R, mRNA sequence.
ACCESSION BQ244558
VERSION BQ244558.1 GI:20440434
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 451)
AUTHORS Cloutier,S.
TITLE Wheat functional genomics - Glenlea developing seeds cDNA libraries
JOURNAL Unpublished (2002)
COMMENT Contact: Dr. Sylvie Cloutier
Cereal Research Centre, Agriculture and Agri-food Canada
195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
Tel: (204) 983-2340
Fax: (204) 983-4604
Email: scloutier@agr.gc.ca
was cloned directionally, not all sequences generated with reverse primer were from the 5' end (same with forward primer and 3' end).
Average insert size is >1.4 kb
Plate: 040 row: H column: 11
Seq primer: M13 Reverse.
FEATURES Location/Qualifiers
source 1. .451
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Glenlea"
/db_xref="taxon:4565"
/clone="TaE15040H11R"
/tissue_type="developing seeds"
/dev_stage="15 days after anthesis"
/lab_host="E. coli DH10B"
/clone_lib="TaE15"
/note="Vector: pCMV-SPORT6.0 (Invitrogen Technologies); Site_1: NotI; Site_2: MluI; mRNA obtained from wheat seeds of cultivar Glenlea 15 days post-anthesis"
ORIGIN

Alignment Scores:
Pred. No.: 2.59e+03 Length: 451
Score: 42.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-09-858-332G-28 (1-6) x BQ244558 (1-451)

Qy 1 HisAsnHisAsnHisAsn 6
||||||||||||||||||
Db 332 CACAACCACAACCACAAT 349

RESULT 379
BQ490899
LOCUS BQ490899 451 bp mRNA linear EST 31-OCT-2002
DEFINITION EST00128 Pb0001 Paracoccidioides brasiliensis cDNA, mRNA sequence.
ACCESSION BQ490899
VERSION BQ490899.1 GI:24433972
KEYWORDS EST.
SOURCE Paracoccidioides brasiliensis
ORGANISM Paracoccidioides brasiliensis
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Onygenales; mitosporic Onygenales; Paracoccidioides.
REFERENCE 1 (bases 1 to 451)
AUTHORS Goldman,G.H., dos Reis Marques,E., Duarte Ribeiro,D.C., de Souza Bernardes,L.A., Quiapin,A.C., Vitorelli,P.M., Savoldi,M., Semighini,C.P., de Oliveira,R.C., Nunes,L.R., Travassos,L.R., Puccia,R., Batista,W.L., Ferreira,L.E., Moreira,J.C., Bogossian,A.P., Tekaia,F., Nobrega,M.P., Nobrega,F.G. and Goldman,M.H.
TITLE Expressed sequence tag analysis of the human pathogen Paracoccidioides brasiliensis yeast phase: identification of putative homologues of Candida albicans virulence and pathogenicity genes
JOURNAL Eukaryot. Cell 2 (1), 34-48 (2003)
PUBMED 12582121
COMMENT Contact: Gustavo Henrique Goldman
Laboratory of Molecular Biology
Universidade de Sao Paulo - USP - FCFRP
Av do Cafe S/N, CEP: 14040-903, Ribeirao Preto - SP, Brazil
Email: ggoldman@usp.br.
FEATURES Location/Qualifiers
source 1. .451
/organism="Paracoccidioides brasiliensis"
/mol_type="mRNA"
/db_xref="taxon:121759"
/clone_lib="Pb0001"
ORIGIN

Alignment Scores:
Pred. No.: 2.59e+03 Length: 451
Score: 42.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 5 Gaps: 0

US-09-858-332G-28 (1-6) x BQ490899 (1-451)

Qy 1 HisAsnHisAsnHisAsn 6
||||||||||||||||||
Db 434 CACAATCACAACCATAAC 451

RESULT 380
C94403
LOCUS C94403 451 bp mRNA linear EST 15-JUN-1998
DEFINITION C94403 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium discoideum cDNA clone SSK652, mRNA sequence.
ACCESSION C94403
VERSION C94403.1 GI:3219018
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 451)
AUTHORS Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M., Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H., Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
TITLE Developmental cDNA in Dictyostelium discoideum
JOURNAL Unpublished (1998)
COMMENT Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba

```
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp
            PROJECT = Dictyostelium discoideum cDNA project in Japan.
FEATURES              Location/Qualifiers
     source           1. .451
                      /organism="Dictyostelium discoideum"
                      /mol_type="mRNA"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /clone="SSK652"
                      /dev_stage="slug"
                      /clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
ORIGIN

Alignment Scores:
Pred. No.:                2.59e+03        Length:         451
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       5               Gaps:           0

US-09-858-332G-28 (1-6) x C94403 (1-451)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        138 CACAACCACAACCACAAC 155

RESULT 381
CB821385/c
LOCUS       CB821385                451 bp    mRNA    linear   EST 16-MAY-2003
DEFINITION  EST 2236 Half-Ripe Apricot Fruit Lambda Zap II Library Prunus
            armeniaca cDNA clone be002e21 3', mRNA sequence.
ACCESSION   CB821385
VERSION     CB821385.1  GI:29954591
KEYWORDS    EST.
SOURCE      Prunus armeniaca (apricot)
  ORGANISM  Prunus armeniaca
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Rosales; Rosaceae; Amygdaloideae; Prunus.
REFERENCE   1  (bases 1 to 451)
  AUTHORS   Grimplet,J., Romieu,C., Audergon,J.M., Albagnac,G., Lambert,P.,
            Bouchet,J.P. and Terrier,N.
  TITLE     High Throughput Detection of Isogenes among 5724 3' EST from
            Apricot Fruit(Prunus armeniaca)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Audergon JM
            Unite de genetique et amelioration des fruits et legumes
            Institut National de la Recherche Agronomique
            Domaine Saint-Maurice BP 94 84143 Montfavet cedex
            Tel: 00-33-(0)4-32-72-26-68
            Fax: 00-33-(0)4-32-72-26-62
            Email: audergon@avignon.inra.fr
            Seq primer: T7.
FEATURES              Location/Qualifiers
     source           1. .451
                      /organism="Prunus armeniaca"
                      /mol_type="mRNA"
                      /cultivar="Bergeron"
                      /db_xref="taxon:36596"
                      /clone="be002e21"
                      /dev_stage="Half-Ripe stage"
                      /clone_lib="Half-Ripe Apricot Fruit Lambda Zap II Library"
                      /note="Organ: Fruit; Vector: Lambda Zap II; Site_1: Eco
                      RI; Site_2: XhoI; Oriented library, construction described
                      in Molecular cloning and expression of a cDNA encoding
                      1-aminocyclopropane-1-carboxylate (ACC) oxidase from
                      apricot fruit (Prunus armeniaca cv. Bergeron) by
                      Mbeguie-Mbeguie D, Chahine H, Gomez RM, Gouble B, Audergon
                      JM, Souty M, Albagnac G, Fils-Lycaon B in Physiol Plant
                      105:294-303 1999"
ORIGIN

Alignment Scores:
Pred. No.:                2.59e+03        Length:         451
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x CB821385 (1-451)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        333 CACAACCACAACCACAAC 316

RESULT 382
H11071
LOCUS       H11071                  451 bp    mRNA    linear   EST 26-JUN-1995
DEFINITION  ym09e08.s1 Soares infant brain 1NIB Homo sapiens cDNA clone
            IMAGE:47114 3', mRNA sequence.
ACCESSION   H11071
VERSION     H11071.1  GI:875891
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 451)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1469
            High quality sequence stops: 356
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1469   Std Error: 0.00
            Seq primer: Promega -21m13
            High quality sequence stop: 356.
FEATURES              Location/Qualifiers
     source           1. .451
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="GDB:419655"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:47114"
                      /sex="female"
                      /dev_stage="73 days post natal"
                      /lab_host="DH10B (ampicillin resistant)"
                      /clone_lib="Soares infant brain 1NIB"
                      /note="Organ: whole brain; Vector: Lafmid BA; Site_1: Not
                      I; Site_2: Hind III; 1st strand cDNA was primed with a Not
                      I - oligo(dT) primer [5'
                      AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTTT 3'];
                      double-stranded cDNA was ligated to Hind III adaptors
                      (Pharmacia), digested with Not I and directionally cloned
                      into the Not I and Hind III sites of the Lafmid BA vector.
                      Library went through one round of normalization. Library
                      constructed by Bento Soares and M.Fatima Bonaldo."
ORIGIN
```

```
Alignment Scores:
Pred. No.:              2.59e+03        Length:         451
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     8               Gaps:           0

US-09-858-332G-28 (1-6) x H11071 (1-451)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        199 CATAATCATAATCATAAT 216

RESULT 383
AI721582
LOCUS       AI721582                452 bp    mRNA    linear   EST 07-JUN-2001
DEFINITION  fc29f06.x1 Zebrafish WashU MPIMG EST Danio rerio cDNA clone
            IMAGE:3722819 3', mRNA sequence.
ACCESSION   AI721582
VERSION     AI721582.1  GI:5039911
KEYWORDS    EST.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 452)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M.,
            Eddy,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
            Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
            Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     Other_ESTs: fc29f06.y1
            Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
            Matthew Clark. DNA Sequencing by: Washington University Genome
            Sequencing Center Clone distribution: Genome Systems, St. Louis,
            Missouri (web address: www.genomesystems.com) (email contact:
            info@genomesystems.com) and Research Genetics, Huntsville, Alabama
            (web address: www.resgen.com) (email contact: info@resgen.com) and
            RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
            www.rzpd.de)
            Seq primer: T7 ET from Amersham
            High quality sequence stop: 433.
FEATURES             Location/Qualifiers
     source          1. .452
                     /organism="Danio rerio"
                     /mol_type="mRNA"
                     /db_xref="taxon:7955"
                     /clone="IMAGE:3722819"
                     /sex="mixed"
                     /tissue_type="26 somite embryos, adult livers, shield
                     stage embryos"
                     /lab_host="XL1-blue MRF"
                     /clone_lib="Zebrafish WashU MPIMG EST"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
                     strand cDNA was primed with a Not I - oligo(dT)15 primer
                     [5'pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTT3']
                     double-stranded cDNA was ligated to Sal I adaptors (BRL),
                     digested with Not I and cloned into the Not I and Sal I
                     sites of the pSPORT1 vector (BRL). Library was constructed
                     by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
                     Institut fuer Molekulare Genetik,Berlin). cDNAs for EST
                     analysis were selected following oligonucleotide
                     hybridization fingerprinting of arrayed clones from
                     zebrafish late somitogenesis (26 ss), adult liver or
                     embryonic shield stage (5.6 h) libraries. Fingerprint
                     data were used to computationally cluster cDNAs, and a
                     single cDNA from each cluster was chosen for sequencing.
                     In some cases multiple members of the same cluster were
                     sequenced to assess clustering parameters or single clones
                     were sequenced additional times to assess quality
                     control."
ORIGIN

Alignment Scores:
Pred. No.:              2.6e+03         Length:         452
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AI721582 (1-452)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        186 CATAATCATAATCATAAT 203

RESULT 384
AU269566
LOCUS       AU269566                452 bp    mRNA    linear   EST 26-APR-2004
DEFINITION  AU269566 VS Dictyostelium discoideum cDNA clone VSJ203 3', mRNA
            sequence.
ACCESSION   AU269566
VERSION     AU269566.1  GI:20528364
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 452)
  AUTHORS   Urushihara,H., Morio,T., Saito,T., Kohara,Y., Koriki,E., Ochiai,H.,
            Maeda,M., Williams,J.G., Takeuchi,I. and Tanaka,Y.
  TITLE     Analyses of cDNAs from growth and slug stages of Dictyostelium
            discoideum
  JOURNAL   Nucleic Acids Res. 32 (5), 1647-1653 (2004)
   PUBMED   15010511
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .452
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="VSJ203"
                     /sex="mat A"
                     /dev_stage="vegetative"
                     /clone_lib="VS"
ORIGIN

Alignment Scores:
Pred. No.:              2.6e+03         Length:         452
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AU269566 (1-452)
```

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        304 CACAACCACAACCACAAC 321

RESULT 385
BP591415/c
LOCUS       BP591415                 452 bp    mRNA    linear   EST 23-JUN-2004
DEFINITION  BP591415 RAFL15 Arabidopsis thaliana cDNA clone RAFL15-14-J18 3',
            mRNA sequence.
ACCESSION   BP591415
VERSION     BP591415.1  GI:49158883
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 452)
  AUTHORS   Seki,M., Narusaka,M., Kamiya,A., Ishida,J., Satou,M., Sakurai,T.,
            Nakajima,M., Enju,A., Akiyama,K., Oono,Y., Muramatsu,M.,
            Hayashizaki,Y., Kawai,J., Carninci,P., Itoh,M., Ishii,Y.,
            Arakawa,T., Shibata,K., Shinagawa,A. and Shinozaki,K.
  TITLE     Functional annotation of a full-length Arabidopsis cDNA collection
  JOURNAL   Science 296 (5565), 141-145 (2002)
   PUBMED   11910074
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            reversed clone; Please visit our web site
            (http://pfgweb.gsc.riken.go.jp/) for further details.
FEATURES             Location/Qualifiers
     source          1. .452
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="RAFL15-14-J18"
                     /tissue_type="mixture of silique and flower"
                     /lab_host="DH10B"
                     /clone_lib="RAFL15"
                     /note="Site_1: BamHI; Site_2: SalI"
ORIGIN

Alignment Scores:
Pred. No.:                2.6e+03        Length:          452
Score:                    42.00          Matches:         6
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       3              Gaps:            0

US-09-858-332G-28 (1-6) x BP591415 (1-452)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        305 CACAACCACAACCACAAC 288

RESULT 386
DN373078/c
LOCUS       DN373078                 452 bp    mRNA    linear   EST 07-MAR-2005
DEFINITION  LIB3733-052-B2-K1-D6 LIB3733 Canis familiaris cDNA clone
            CLN12921935, mRNA sequence.
ACCESSION   DN373078
VERSION     DN373078.1  GI:60554298
KEYWORDS    EST.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
            Canis.
REFERENCE   1  (bases 1 to 452)
  AUTHORS   Staten,N.R.
  TITLE     Direct Submission (Staten,N.R.)
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Nick Staten
            Tel: 636 247 6855
            Email: nicholas.r.staten@pfizer.com.
FEATURES             Location/Qualifiers
     source          1. .452
                     /organism="Canis familiaris"
                     /mol_type="mRNA"
                     /strain="beagle"
                     /db_xref="taxon:9615"
                     /clone="CLN12921935"
                     /tissue_type="brain"
                     /lab_host="BHK"
                     /clone_lib="LIB3733"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; brain
                     - dog"
ORIGIN

Alignment Scores:
Pred. No.:                2.6e+03        Length:          452
Score:                    42.00          Matches:         6
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       8              Gaps:            0

US-09-858-332G-28 (1-6) x DN373078 (1-452)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         40 CATAATCATAATCATAAT 23

RESULT 387
AU040068
LOCUS       AU040068                 453 bp    mRNA    linear   EST 29-MAR-1999
DEFINITION  AU040068 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
            discoideum cDNA clone SLA355, mRNA sequence.
ACCESSION   AU040068
VERSION     AU040068.1  GI:4009308
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 453)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     The Dictyostelium developmental cDNA project: generation and
            analysis of expressed sequence tags from the first-finger stage of
            development
  JOURNAL   DNA Res. 5 (6), 335-340 (1998)
   PUBMED   10048482
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp
            PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1. .453
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SLA355"
                     /dev_stage="slug"
                     /clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
ORIGIN
```

```
Alignment Scores:
Pred. No.:              2.61e+03        Length:         453
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AU040068 (1-453)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        233 CACAATCACAATCACAAT 250

RESULT 388
BP595105/c
LOCUS       BP595105                  453 bp    mRNA    linear   EST 23-JUN-2004
DEFINITION  BP595105 RAFL15 Arabidopsis thaliana cDNA clone RAFL15-28-I10 3',
            mRNA sequence.
ACCESSION   BP595105
VERSION     BP595105.1  GI:49162573
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 453)
  AUTHORS   Seki,M., Narusaka,M., Kamiya,A., Ishida,J., Satou,M., Sakurai,T.,
            Nakajima,M., Enju,A., Akiyama,K., Oono,Y., Muramatsu,M.,
            Hayashizaki,Y., Kawai,J., Carninci,P., Itoh,M., Ishii,Y.,
            Arakawa,T., Shibata,K., Shinagawa,A. and Shinozaki,K.
  TITLE     Functional annotation of a full-length Arabidopsis cDNA collection
  JOURNAL   Science 296 (5565), 141-145 (2002)
   PUBMED   11910074
COMMENT     Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            reversed clone; Please visit our web site
            (http://pfgweb.gsc.riken.go.jp/) for further details.
FEATURES             Location/Qualifiers
     source          1. .453
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="RAFL15-28-I10"
                     /tissue_type="mixture of silique and flower"
                     /lab_host="DH10B"
                     /clone_lib="RAFL15"
                     /note="Site_1: BamHI; Site_2: SalI"
ORIGIN

Alignment Scores:
Pred. No.:              2.61e+03        Length:         453
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-858-332G-28 (1-6) x BP595105 (1-453)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        306 CACAACCACAACCACAAC 289

RESULT 389
CN193518/c
LOCUS       CN193518                  453 bp    mRNA    linear   EST 05-APR-2004
DEFINITION  rg88a10.y1 Meloidogyne hapla female SMART pGEM Meloidogyne hapla
            cDNA 5', mRNA sequence.
ACCESSION   CN193518
VERSION     CN193518.1  GI:46218457
KEYWORDS    EST.
SOURCE      Meloidogyne hapla
  ORGANISM  Meloidogyne hapla
            Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
            Tylenchoidea; Meloidogynidae; Meloidogyninae; Meloidogyne.
REFERENCE   1  (bases 1 to 453)
  AUTHORS   McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J.,
            Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B.,
            Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C.,
            Tsagareishvili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C.,
            Underwood,K., Steptoe,M., Allen,M., Person,B., Swaller,T.,
            Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     The Washington Univ. Nematode EST Project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: McCarter JP
            The Washington Univ. Nematode EST Project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Cloned unidirectionally. Poly(A)+ RNA was concentrated and purified
            using Dynabeads (Dynal) and mRNA eluted for first strand synthesis.
            First strand cDNA was created using MMLV RT (Powerscript, Clontech)
            and primed with oligo(dT) with XhoI site and 5'SMART 'anchor' added
            using chimeric DNA-RNA oligo. 12 PCR cycles were done using first
            strand and primers specific to SMART oligo and 3' end. Double
            stranded cDNA was digested using XhoI/NotI, fractioned on
            Chroma-spin 400 columns (Clontech) and ligated to digested
            pGEM-11zf(+) plasmid. Chemically competent DH10B cells were used as
            host cells. Library materials provided by Dr. David Bird of North
            Carolina State University. Library construction by Jeff Rousch. See
            www.nematode.net for additional project information.
            Putative full length read
            The vector to vector length is 469
            Seq primer: Sp6.
FEATURES             Location/Qualifiers
     source          1. .453
                     /organism="Meloidogyne hapla"
                     /mol_type="mRNA"
                     /db_xref="taxon:6305"
                     /tissue_type="whole organism"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="Meloidogyne hapla female SMART pGEM"
                     /note="Vector: plasmid (ampicillin resistant); Site_1:
                     XhoI; Site_2: NotI; Cloned unidirectionally. Poly(A)+ RNA
                     was concentrated and purified using Dynabeads (Dynal) and
                     mRNA eluted for first strand synthesis. First strand cDNA
                     was created using MMLV RT (Powerscript, Clontech) and
                     primed with oligo(dT) with XhoI site and 5'SMART 'anchor'
                     added using chimeric DNA-RNA oligo. 12 PCR cycles were
                     done using first strand and primers specific to SMART
                     oligo and 3' end. Double stranded cDNA was digested using
                     XhoI/NotI, fractioned on Chroma-spin 400 columns
                     (Clontech) and ligated to digested pGEM-11zf(+) plasmid.
                     Chemically competent DH10B cells were used as host cells.
                     Library materials provided by Dr. David Bird of North
                     Carolina State University. Library construction by Jeff
                     Rousch. See www.nematode.net for additional project
                     information."
ORIGIN

Alignment Scores:
Pred. No.:              2.61e+03        Length:         453
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
```

```
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     7              Gaps:          0

US-09-858-332G-28 (1-6) x CN193518 (1-453)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        259 CATAATCACAATCATAAT 242

RESULT 390
CN242783
LOCUS       CN242783                454 bp    mRNA    linear   EST 09-APR-2004
DEFINITION  EST008655 Mycelium and yeast cells from Paracoccidioides
            brasiliensis Paracoccidioides brasiliensis cDNA, mRNA sequence.
ACCESSION   CN242783
VERSION     CN242783.1  GI:46346527
KEYWORDS    EST.
SOURCE      Paracoccidioides brasiliensis
  ORGANISM  Paracoccidioides brasiliensis
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Onygenales; mitosporic Onygenales; Paracoccidioides.
REFERENCE   1  (bases 1 to 454)
  AUTHORS   Felipe,M.S., Andrade,R.V., Arraes,F.B., Nicola,A.M., Maranhao,A.Q.,
            Torres,F.A., Silva-Pereira,I., Pocas-Fonseca,M.J., Campos,E.G.,
            Moraes,L.M., Andrade,P.A., Tavares,A.H., Silva,S.S., Kyaw,C.M.,
            Souza,D.P., Network,P., Pereira,M., Jesuino,R.S., Andrade,E.V.,
            Parente,J.A., Oliveira,G.S., Barbosa,M.S., Martins,N.F.,
            Fachin,A.L., Cardoso,R.S., Passos,G.A., Almeida,N.F., Walter,M.E.,
            Soares,C.M., Carvalho,M.J. and Brigido,M.M.
  TITLE     Transcriptional Profiles of the Human Pathogenic Fungus
            Paracoccidioides brasiliensis in Mycelium and Yeast Cells
  JOURNAL   J. Biol. Chem. 280 (26), 24706-24714 (2005)
  PUBMED    15849188
COMMENT     Contact: Felipe MSS
            Laboratory of Molecular Biology
            Institute of Biology - University of Brasilia
            Campus Universitario, Asa Norte, Brasilia, DF 70910-900, BRA
            Tel: 55 61 307 2423
            Fax: 55 61 349 8411
            Email: msueli@unb.br
            Seq primer: T7 Sequencing primer.
FEATURES             Location/Qualifiers
     source          1. .454
                     /organism="Paracoccidioides brasiliensis"
                     /mol_type="mRNA"
                     /strain="Pb01"
                     /db_xref="taxon:121759"
                     /clone_lib="Mycelium and yeast cells from Paracoccidioides
                     brasiliensis"
                     /note="Pb Lambda Zap Express Library"
ORIGIN

Alignment Scores:
Pred. No.:              2.61e+03       Length:        454
Score:                  42.00          Matches:       6
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     7              Gaps:          0

US-09-858-332G-28 (1-6) x CN242783 (1-454)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        167 CATAACCATAACCATAAC 184

RESULT 391
CN828903/c
LOCUS       CN828903                454 bp    mRNA    linear   EST 14-MAY-2005
DEFINITION  EL3189R Brassica embryo library (EL) Brassica napus cDNA clone
            EL3189 5', mRNA sequence.
ACCESSION   CN828903
VERSION     CN828903.1  GI:65298689
KEYWORDS    EST.
SOURCE      Brassica napus (rape)
  ORGANISM  Brassica napus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 454)
  AUTHORS   Sharpe,A.G., Gjetvaj,B., Durkin,J. and Lydiate,D.J.
  TITLE     Brassica napus ESTs
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Sharpe, A.G.
            Molecular Genetics
            Agriculture & Agri-Food Canada
            107 Science Place, Saskatoon, Sakatchewan, Canada, S7N0X2
            Tel: 306 956 7271
            Fax: 306 956 7247
            Email: sharpea@agr.gc.ca
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .454
                     /organism="Brassica napus"
                     /mol_type="mRNA"
                     /cultivar="DH12075 (double haploid line from Cresor x
                     Westar cross)"
                     /db_xref="taxon:3708"
                     /clone="EL3189"
                     /dev_stage="Mid to late embryos (4-6 mg)"
                     /lab_host="E. coli ElectroMAX DH5 alpha-e (Invitrogen)"
                     /clone_lib="Brassica embryo library (EL)"
                     /note="Organ: Embryos without seed coat; Vector: pSPORT1
                     (modified: GCGGCCGCCc*GACTAGTGAGCTC*cggacgcgtggGTCGAC);
                     Site_1: NotI; Site_2: SalI; Seeds were collected by Dr.
                     Francois Ouellet when they were still very green (mid to
                     large stage, cotyledons were formed). The seed coats were
                     removed and the remaining tissue was used for cDNA library
                     construction. mRNA was poly-A primed using SuperScript
                     Plasmid System cDNA Synthesis and Cloning kit (Invitrogen)
                     After initial screening, the most abundant redundant
                     clones were screened out using 22 oligos designed to match
                     napins (including albumins), cruciferins, oleosins,
                     trypsin inhibitor 2, cytosolic GAPDH, cyclophilins, HSP70,
                     desaturase, and CAB (LHCP)."
ORIGIN

Alignment Scores:
Pred. No.:              2.61e+03       Length:        454
Score:                  42.00          Matches:       6
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     7              Gaps:          0

US-09-858-332G-28 (1-6) x CN828903 (1-454)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        135 CATAACCATAACCATAAC 118

RESULT 392
W50496
LOCUS       W50496                  454 bp    mRNA    linear   EST 29-MAY-1996
DEFINITION  md04g05.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:367448 5' similar to gb:M38188 OVARIAN GRANULOSA CELL
            13.0 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION   W50496
VERSION     W50496.1  GI:1339137
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
```

```
REFERENCE   1  (bases 1 to 454)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:228880
            Seq primer: mob.REGA+ET
            High quality sequence stop: 352.
FEATURES             Location/Qualifiers
     source          1. .454
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:367448"
                     /sex="unknown"
                     /tissue_type="embryo"
                     /dev_stage="13.5-14.5dpc total fetus"
                     /lab_host="DH10B"
                     /clone_lib="Soares mouse embryo NbME13.5 14.5"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTTT
                     T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                     14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                     State Univ., from 2 ]; double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector. Library went through one round of
                     normalization, and was constructed by Bento Soares and
                     M.Fatima Bonaldo.  "
ORIGIN

Alignment Scores:
Pred. No.:                2.61e+03        Length:         454
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       8               Gaps:           0

US-09-858-332G-28 (1-6) x W50496 (1-454)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        285 CACAACCATAACCACAAC 302

RESULT 393
AI183075
LOCUS       AI183075                455 bp    mRNA    linear   EST 08-OCT-1998
DEFINITION  ub93e07.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
            IMAGE:1396068 5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0
            KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION   AI183075
VERSION     AI183075.1  GI:3733713
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 455)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:907784
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 450.
FEATURES             Location/Qualifiers
     source          1. .455
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1396068"
                     /sex="male"
                     /tissue_type="mammary gland"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /clone_lib="Soares_mammary_gland_NbMMG"
                     /note="Organ: mammary gland; Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
ORIGIN

Alignment Scores:
Pred. No.:                2.62e+03        Length:         455
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       1               Gaps:           0

US-09-858-332G-28 (1-6) x AI183075 (1-455)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        104 CACAACCATAACCACAAC 121

RESULT 394
BJ432539/c
LOCUS       BJ432539                455 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ432539 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv19a06 3', mRNA sequence.
ACCESSION   BJ432539
VERSION     BJ432539.1  GI:19407261
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 455)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
```

```
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .455
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv19a06"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:                2.62e+03        Length:         455
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       3               Gaps:           0

US-09-858-332G-28 (1-6) x BJ432539 (1-455)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        429 CACAATCACAATCACAAT 412

RESULT 395
AA444856
LOCUS       AA444856                456 bp    mRNA    linear   EST 03-JUN-1997
DEFINITION  vg57c09.r1 Beddington mouse embryonic region Mus musculus cDNA
            clone IMAGE:865456 5' similar to gb:M38188 OVARIAN GRANULOSA CELL
            13.0 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION   AA444856
VERSION     AA444856.1  GI:2157633
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 456)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:509544
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 250.
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6 x DBA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:865456"
                     /sex="pooled"
                     /tissue_type="embryo"
                     /dev_stage="7.5dpc"
                     /lab_host="DH12S"
                     /clone_lib="Beddington mouse embryonic region"
                     /note="Organ: whole embryo; Vector: pCMV-SPORT; Site_1:
                     SalI; Site_2: NotI; Cloned unidirectionally.  Primer:
                     Oligo dT. Gastrulating embryos were collected at 7.5dpc
                     from C57BL6 x DBA matings, excluding embryos that had
                     developed head folds and all extraembryonic tissues.
                     Average insert size: 1.3 kb (range: 0.5 - 3.0 kb).
                     Referenced in Development 121, 2479-2489 (1995)"
ORIGIN

Alignment Scores:
Pred. No.:                2.62e+03        Length:         456
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       1               Gaps:           0

US-09-858-332G-28 (1-6) x AA444856 (1-456)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        247 CACAACCATAACCACAAC 264

RESULT 396
BF483293
LOCUS       BF483293                456 bp    mRNA    linear   EST 06-DEC-2000
DEFINITION  WHE1791_B05_C09ZS Wheat pre-anthesis spike cDNA library Triticum
            aestivum cDNA clone WHE1791_B05_C09, mRNA sequence.
ACCESSION   BF483293
VERSION     BF483293.1  GI:11566594
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 456)
  AUTHORS   Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D.,
            Han,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J.,
            Seaton,C.L. and Tong,J.C.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Pre-anthesis spike cDNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene SK primer.
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /cultivar="Chinese Spring"
                     /db_xref="taxon:4565"
                     /clone="WHE1791_B05_C09"
                     /tissue_type="Spike before anthesis"
                     /dev_stage="Adult plant"
                     /lab_host="E. coli SOLR"
                     /clone_lib="Wheat pre-anthesis spike cDNA library"
```

/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site_1: EcoRI; Site_2: XhoI; Plants were grown in the greenhouse. Whole spike with awns trimmed, white, green and yellow anther were collected and total RNA, and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give pBluescript phagemids in the TJ Close lab (Choi, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."
ORIGIN

Alignment Scores:
Pred. No.: 2.62e+03 Length: 456
Score: 42.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 2 Gaps: 0

US-09-858-332G-28 (1-6) x BF483293 (1-456)

Qy 1 HisAsnHisAsnHisAsn 6
||||||||||||||||||
Db 161 CACAACCACAACCACAAT 178

RESULT 397
CF985606
LOCUS CF985606 456 bp mRNA linear EST 04-FEB-2005
DEFINITION 24107rsicee_5733.y1 Oryza sativa cv. LYP9 tillering whole plant cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA sequence.
ACCESSION CF985606
VERSION CF985606.1 GI:58587298
KEYWORDS EST.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 456)
AUTHORS Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W., Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X., Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J., Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q., Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z., Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P., Wu,S. and Liu,J.
TITLE The Genomes of Oryza sativa: A History of Duplications
JOURNAL PLoS Biol. 3 (2), e38 (2005)
PUBMED 15685292
COMMENT Contact: Yan Zhou
Bioinformatics Department
Hangzhou Genomics Institute
No.51 Zhijiang Road, Hangzhou 310008, China
Tel: 86-571-56805886
Fax: 86-571-56805884
Email: zhouyan@genomics.org.cn
Seq primer: M13 Forward
High quality sequence stop: 456
POLYA=No.
FEATURES Location/Qualifiers
source 1. .456
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/cultivar="LYP9"
/db_xref="taxon:39946"
/tissue_type="whole plant"
/dev_stage="tillering"
/clone_lib="Oryza sativa cv. LYP9 tillering whole plant cDNA library"
ORIGIN

Alignment Scores:
Pred. No.: 2.62e+03 Length: 456
Score: 42.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0

US-09-858-332G-28 (1-6) x CF985606 (1-456)

Qy 1 HisAsnHisAsnHisAsn 6
||||||||||||||||||
Db 203 CACAACCACAACCACAAC 220

RESULT 398
BE654418
LOCUS BE654418 457 bp mRNA linear EST 06-SEP-2000
DEFINITION UI-M-AH1-agu-b-10-0-UI.r1 NIH_BMAP_MCE_N Mus musculus cDNA clone UI-M-AH1-agu-b-10-0-UI 5', mRNA sequence.
ACCESSION BE654418
VERSION BE654418.1 GI:9980331
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 457)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
PUBMED 8889548
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
cDNA Library Preparation: M.B. Soares Lab Clone distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements The following repetitive elements were found in this cDNA sequence: 263-304, >(CAA)n#Simple_repeat
Seq primer: M13 Reverse.
FEATURES Location/Qualifiers
source 1. .457
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-AH1-agu-b-10-0-UI"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NIH_BMAP_MCE_N"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The NIH_BMAP_MCE library is a normalized library constructed from mouse cerebellum. The tag is a string of 5 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996. Tissue provided by Ms. Annie Novakovich, Zivic-Miller Laboratories."
ORIGIN

Alignment Scores:
Pred. No.: 2.63e+03 Length: 457
Score: 42.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0

```
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     2              Gaps:          0

US-09-858-332G-28 (1-6) x BE654418 (1-457)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        281 CACAACCATAACCACAAC 298

RESULT 399
C90662
LOCUS       C90662                   457 bp    mRNA    linear   EST 20-APR-1998
DEFINITION  C90662 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
            discoideum cDNA clone SSI887, mRNA sequence.
ACCESSION   C90662
VERSION     C90662.1  GI:3060282
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 457)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .457
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SSI887"
                     /dev_stage="slug"
                     /clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
ORIGIN

Alignment Scores:
Pred. No.:              2.63e+03       Length:        457
Score:                  42.00          Matches:       6
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     5              Gaps:          0

US-09-858-332G-28 (1-6) x C90662 (1-457)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        230 CATAATCATAATCATAAT 247

RESULT 400
CF985050
LOCUS       CF985050                 457 bp    mRNA    linear   EST 04-FEB-2005
DEFINITION  17245rsicee_15083.y1 Oryza sativa cv. LYP9 tillering whole plant
            cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF985050
VERSION     CF985050.1  GI:58586742
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 457)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
  PUBMED    15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 457
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .457
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="tillering"
                     /clone_lib="Oryza sativa cv. LYP9 tillering whole plant
                     cDNA library"
ORIGIN

Alignment Scores:
Pred. No.:              2.63e+03       Length:        457
Score:                  42.00          Matches:       6
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     7              Gaps:          0

US-09-858-332G-28 (1-6) x CF985050 (1-457)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        214 CACAACCACAACCACAAC 231

RESULT 401
AA184544
LOCUS       AA184544                 458 bp    mRNA    linear   EST 12-FEB-1997
DEFINITION  mt50h01.r1 Stratagene mouse embryonic carcinoma (#937317) Mus
            musculus cDNA clone IMAGE:633361 5' similar to gb:M38188 OVARIAN
            GRANULOSA CELL 13.0 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION   AA184544
VERSION     AA184544.1  GI:1768194
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 458)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
```

```
           Tel: 314 286 1800
           Fax: 314 286 1810
           Email: mouseest@watson.wustl.edu
           This clone is available royalty-free through LLNL ; contact the
           IMAGE Consortium (info@image.llnl.gov) for further information.
           MGI:385353
           Seq primer: -28m13 rev1 ET from Amersham
           High quality sequence stop: 341.
FEATURES             Location/Qualifiers
     source          1. .458
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:633361"
                     /tissue_type="carcinoma"
                     /dev_stage="embryonic"
                     /lab_host="SOLR (kanamycin resistant)"
                     /clone_lib="Stratagene mouse embryonic carcinoma
                     (#937317)"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; Cloned unidirectionally.  Primer: Oligo dT. P19 cell
                     line. Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5'
                     adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
ORIGIN

Alignment Scores:
Pred. No.:              2.64e+03       Length:         458
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     1              Gaps:           0

US-09-858-332G-28 (1-6) x AA184544 (1-458)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        203 CACAACCATAACCACAAC 220

RESULT 402
AL368151
LOCUS       AL368151                458 bp    mRNA    linear   EST 03-AUG-2000
DEFINITION  MtBA22E07F1 MtBA Medicago truncatula cDNA clone MtBA22E07 T3, mRNA
            sequence.
ACCESSION   AL368151
VERSION     AL368151.1  GI:9667904
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 458)
  AUTHORS   Journet,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jaillon,O.,
            Niebel,A., Carreau,V., Chatagnier,O., Kahn,D.,
            Gianinazzi-Pearson,V. and Gamas,P.
  TITLE     Medicago truncatula ESTs from nitrogen-starved roots
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de
            Biologie Moleculaire des Relations Plantes-Microorganismes,
            CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :
            Mt-est@toulouse.inra.fr Website :
            http://sequence.toulouse.inra.fr/Mtruncatula.html).
FEATURES             Location/Qualifiers
     source          1. .458
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /cultivar="Jemalong"
                     /db_xref="taxon:3880"
                     /clone="MtBA22E07"
                     /tissue_type="root tips"
                     /dev_stage="harvested after 3 days of N-starvation"
                     /clone_lib="MtBA"
                     /note="Vector: pBluescript pSK; Site_1: EcoRI; Site_2:
                     XhoI; Plants were grown in an aeroponic chamber for 14
                     days on nitrogen-rich medium followed by 3 days on N-free
                     medium. RNA was extracted from root tips (1-3 cm). cDNA
                     was prepared from polyA+ enriched RNA. The cDNA was
                     directionally ligated into Uni-zapXR vector from
                     Stratagene and packaged using Gigapack Gold packaging
                     extracts. Plasmids containing cDNA inserts were
                     mass-excised from phage stocks using ExAssit helper phage
                     and propagated in SOLR cells. Clone ordering and
                     sequencing was performed by the Centre National de
                     Sequencage (Genoscope, Evry, France)."
ORIGIN

Alignment Scores:
Pred. No.:              2.64e+03       Length:         458
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     1              Gaps:           0

US-09-858-332G-28 (1-6) x AL368151 (1-458)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         46 CACAACCATAACCATAAC 63

RESULT 403
BQ246771
LOCUS       BQ246771                458 bp    mRNA    linear   EST 03-MAY-2002
DEFINITION  TaE15005F12R TaE15 Triticum aestivum cDNA clone TaE15005F12R, mRNA
            sequence.
ACCESSION   BQ246771
VERSION     BQ246771.1  GI:20442647
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 458)
  AUTHORS   Cloutier,S.
  TITLE     Wheat functional genomics - Glenlea developing seeds cDNA libraries
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Dr. Sylvie Cloutier
            Cereal Research Centre, Agriculture and Agri-food Canada
            195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
            Tel: (204) 983-2340
            Fax: (204) 983-4604
            Email: scloutier@agr.gc.ca
            was cloned directionally, not all sequences generated with reverse
            primer were from the 5' end (same with forward primer and 3' end).
            Average insert size is >1.4 kb
            Plate: 005  row: F  column: 12
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .458
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /cultivar="Glenlea"
                     /db_xref="taxon:4565"
                     /clone="TaE15005F12R"
                     /tissue_type="developing seeds"
                     /dev_stage="15 days after anthesis"
                     /lab_host="E. coli DH10B"
                     /clone_lib="TaE15"
```

/note="Vector: pCMV-SPORT6.0 (Invitrogen Technologies); Site_1: NotI; Site_2: MluI; mRNA obtained from wheat seeds of cultivar Glenlea 15 days post-anthesis"
ORIGIN

Alignment Scores:

Pred. No.:	2.64e+03	Length:	458
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	3	Gaps:	0

US-09-858-332G-28 (1-6) x BQ246771 (1-458)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        337 CACAACCACAACCACAAT 354
```

RESULT 404
CA667753
LOCUS CA667753 458 bp mRNA linear EST 24-NOV-2002
DEFINITION wlsu1.pk017.h15 wlsu1 Triticum aestivum cDNA clone wlsu1.pk017.h15 5' end, mRNA sequence.
ACCESSION CA667753
VERSION CA667753.1 GI:25246406
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 458)
AUTHORS Tingey,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z., Miao,G., Caraher,N. and Hanafey,M.K.
TITLE DuPont Wheat cDNA Sequence
JOURNAL Unpublished (2002)
COMMENT Contact: Scott V. Tingey
Crop Genetics
E. I. DuPont de Nemours and Company
1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
Tel: 302-631-2602
Fax: 302-631-2607
Email: Scott.V.Tingey@USA.dupont.com
Seq primer: M13.
FEATURES Location/Qualifiers
source 1. .458
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Stephens"
/db_xref="taxon:4565"
/clone="wlsu1.pk017.h15"
/tissue_type="leaf"
/clone_lib="wlsu1"
/note="Vector: PGEM-T Easy; Wheat (Triticum aestivum L.) WLMK8 cDNAs subtracted with WLM0 cDNAs"
ORIGIN

Alignment Scores:

Pred. No.:	2.64e+03	Length:	458
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-09-858-332G-28 (1-6) x CA667753 (1-458)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        309 CATAATCATAATCATAAC 326
```

RESULT 405
CO312414/c
LOCUS CO312414 458 bp mRNA linear EST 28-JUN-2004
DEFINITION EK228707.5prime Exelixis FlyTag CK01 pCDNA-SK+ Drosophila melanogaster cDNA clone EK228707 5, mRNA sequence.
ACCESSION CO312414
VERSION CO312414.1 GI:49367809
KEYWORDS EST.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 458)
AUTHORS Kopczynski,C., Platt,D., Campbell,J., Muzong,C., Laufer,A., Peterson,E. and Swimmer,C.
TITLE Exelixis FlyTag EST Project CK01 Library
JOURNAL Unpublished (2004)
COMMENT Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: EK.2287 row: A column: 7
High quality sequence stop: 455.
FEATURES Location/Qualifiers
source 1. .458
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="EK228707"
/clone_lib="Exelixis FlyTag CK01 pCDNA-SK+"
/note="Organ: mixed stage embryos, imaginal disks, and adult heads; Vector: pCDNA-SK+; Site_1: NotI; Site_2: XhoI; Random primed, normalized library from mixed stage embryos, imaginal disks, and adult heads."
ORIGIN

Alignment Scores:

Pred. No.:	2.64e+03	Length:	458
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	7	Gaps:	0

US-09-858-332G-28 (1-6) x CO312414 (1-458)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        370 CATAATCATAATCATAAT 353
```

RESULT 406
AU036977
LOCUS AU036977 459 bp mRNA linear EST 29-MAR-1999
DEFINITION AU036977 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium discoideum cDNA clone SSB896, mRNA sequence.
ACCESSION AU036977
VERSION AU036977.1 GI:3983730
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 459)
AUTHORS Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M., Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H., Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
TITLE The Dictyostelium developmental cDNA project: generation and analysis of expressed sequence tags from the first-finger stage of development
JOURNAL DNA Res. 5 (6), 335-340 (1998)
PUBMED 10048482
COMMENT Contact: Hideko Urushihara

```
           Institute of Biological Sciences
           University of Tsukuba
           1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
           Tel: 81-298-53-4664
           Fax: 81-298-53-6614
           Email: hideko@biol.tsukuba.ac.jp
           PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1. .459
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SSB896"
                     /dev_stage="slug"
                     /clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
ORIGIN

Alignment Scores:
Pred. No.:               2.64e+03       Length:          459
Score:                   42.00          Matches:         6
Percent Similarity:      100.00%        Conservative:    0
Best Local Similarity:   100.00%        Mismatches:      0
Query Match:             100.00%        Indels:          0
DB:                      1              Gaps:            0

US-09-858-332G-28 (1-6) x AU036977 (1-459)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        237 CACAATCACAATCACAAT 254

RESULT 407
BE198421
LOCUS       BE198421                459 bp    mRNA    linear   EST 26-JUN-2000
DEFINITION  ug78d07.y1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
            IMAGE:1548493 5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0
            KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION   BE198421
VERSION     BE198421.1  GI:8710590
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 459)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:953841
            Seq primer: -40RP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .459
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1548493"
                     /sex="female (lactating)"
                     /tissue_type="mammary gland"
                     /lab_host="DH10B"
                     /clone_lib="Soares_mammary_gland_NMLMG"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from mammary
                     gland tissue from a lactating female, and was then primed
                     with a Not I - oligo(dT) primer. Double-stranded  cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I  and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector.  Library is normalized.  Library
                     was constructed by Bento Soares and M. Fatima Bonaldo. "
ORIGIN

Alignment Scores:
Pred. No.:               2.64e+03       Length:          459
Score:                   42.00          Matches:         6
Percent Similarity:      100.00%        Conservative:    0
Best Local Similarity:   100.00%        Mismatches:      0
Query Match:             100.00%        Indels:          0
DB:                      2              Gaps:            0

US-09-858-332G-28 (1-6) x BE198421 (1-459)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         95 CACAACCATAACCACAAC 112

RESULT 408
BM675713
LOCUS       BM675713                460 bp    mRNA    linear   EST 27-FEB-2002
DEFINITION  UI-E-EJ1-ajg-o-24-0-UI.s1 UI-E-EJ1 Homo sapiens cDNA clone
            UI-E-EJ1-ajg-o-24-0-UI 3', mRNA sequence.
ACCESSION   BM675713
VERSION     BM675713.1  GI:18985611
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 460)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
   PUBMED   8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road , 4156  MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            Tissue Procurement: Dr. Gregg Hageman
             cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa
             Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .460
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-E-EJ1-ajg-o-24-0-UI"
                     /tissue_type="fetal eyes, lens, eye anterior segment,
                     optic nerve, retina, Retina Foveal and Macular, RPE and
                     Choroid"
                     /dev_stage="fetal and adult"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /clone_lib="UI-E-EJ1"
                     /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-E-EJ1 is a subtracted cDNA library constructed
                     according to Bonaldo, Lennon and Soares, Genome Research,
                     6:791-806, 1996. First strand cDNA synthesis was primed
                     with an oligo-dT primer containing a Not I site. Double
                     stranded cDNA was ligated to an EcoR I adaptor, digested
                     with Not I, and cloned directionally into pT7T3-Pac
                     vector. The oligonucleotide used to prime the synthesis of
```

first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tags for this library are: fetal eyes, AGAATCAAGA; lens, CGATTAGCGA; eye anterior segment, AATGCCGCAT; optic nerve, CCATTAAGTG; retina, CCGCG; Retina Foveal and Macular, GTCC; RPE and Choroid, ACCTA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI).
TAG_TISSUE=Foveal and Macular Retina
TAG_LIB=UI-E-EJ1
TAG_SEQ=GTCC"

ORIGIN

Alignment Scores:

Pred. No.:	2.65e+03	Length:	460
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	3	Gaps:	0

US-09-858-332G-28 (1-6) x BM675713 (1-460)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        209 CATAATCATAATCATAAT 226
```

RESULT 409
BQ492753
LOCUS BQ492753 460 bp mRNA linear EST 31-OCT-2002
DEFINITION EST01919 Pb0001 Paracoccidioides brasiliensis cDNA, mRNA sequence.
ACCESSION BQ492753
VERSION BQ492753.1 GI:24438733
KEYWORDS EST.
SOURCE Paracoccidioides brasiliensis
ORGANISM Paracoccidioides brasiliensis
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Onygenales; mitosporic Onygenales; Paracoccidioides.
REFERENCE 1 (bases 1 to 460)
AUTHORS Goldman,G.H., dos Reis Marques,E., Duarte Ribeiro,D.C., de Souza Bernardes,L.A., Quiapin,A.C., Vitorelli,P.M., Savoldi,M., Semighini,C.P., de Oliveira,R.C., Nunes,L.R., Travassos,L.R., Puccia,R., Batista,W.L., Ferreira,L.E., Moreira,J.C., Bogossian,A.P., Tekaia,F., Nobrega,M.P., Nobrega,F.G. and Goldman,M.H.
TITLE Expressed sequence tag analysis of the human pathogen Paracoccidioides brasiliensis yeast phase: identification of putative homologues of Candida albicans virulence and pathogenicity genes
JOURNAL Eukaryot. Cell 2 (1), 34-48 (2003)
PUBMED 12582121
COMMENT Contact: Gustavo Henrique Goldman
Laboratory of Molecular Biology
Universidade de Sao Paulo - USP - FCFRP
Av do Cafe S/N, CEP: 14040-903, Ribeirao Preto - SP, Brazil
Email: ggoldman@usp.br.
FEATURES Location/Qualifiers
source 1. .460
/organism="Paracoccidioides brasiliensis"
/mol_type="mRNA"
/db_xref="taxon:121759"
/clone_lib="Pb0001"
ORIGIN

Alignment Scores:

Pred. No.:	2.65e+03	Length:	460
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	5	Gaps:	0

US-09-858-332G-28 (1-6) x BQ492753 (1-460)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db          1 CACAATCACAACCATAAC 18
```

RESULT 410
CF197290
LOCUS CF197290 460 bp mRNA linear EST 31-JUL-2003
DEFINITION Ir64vi2.1_C07 IR64-VI Oryza sativa (indica cultivar-group) cDNA clone IR64-VI2.1C07 5', mRNA sequence.
ACCESSION CF197290
VERSION CF197290.1 GI:33381947
KEYWORDS EST.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 460)
AUTHORS Ventelon-Debout,M., Thi-Thu Huang,N., Wissocq,A., Berger,C., Michele,L., Piegu,B., Cooke,R., Ghesquiere,A., Delseny,M. and Brugidou,C.
TITLE Metabolism and Photosynthesis gene expression is affected in response to rice yellow mottle virus infection in Oryza sativa indica and japonica culltivars
JOURNAL Mol. Genet. Genomics (2003) In press
COMMENT Contact: Brugidou C
IRD
BP 64501, 34394 Montpellier cedex 5 France
Tel: 33 4 67 41 67 39
Fax: 33 4 67 41 61 81
PCR PRimers
FORWARD: 5'AATTAACCCTCACTAAAGGG3'
BACKWARD: 5'CGGGATATCACTCAGCATAATG3'
Plate: 2.1 row: C column: 7.
FEATURES Location/Qualifiers
source 1. .460
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/cultivar="IR64"
/db_xref="taxon:39946"
/clone="IR64-VI2.1C07"
/tissue_type="leaves"
/dev_stage="3 weeks-old plants"
/lab_host="SOLR strain (Stratagene)"
/clone_lib="IR64-VI"
/note="Vector: pBluescript SK (+/-); Site_1: Eco RI; Site_2: Xho I; Sample name: IR64-VI; Plant growth place: greenhouse; Soil conditions: standard; Sowing date: 28/02/2000; Harvesting date: 17/03/2000; Stress date: 15/03/2000; Cultivar: ir64; Description: IR64-RYMV inoculated leaves cDNA library; experimental condition: rymv inoculation;"
ORIGIN

Alignment Scores:

Pred. No.:	2.65e+03	Length:	460
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-09-858-332G-28 (1-6) x CF197290 (1-460)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        202 CACAACCACAACCACAAC 219
```

RESULT 411
BH017468
LOCUS BH017468 460 bp DNA linear GSS 18-MAY-2001
DEFINITION TDGCZ85TH cTOG Lycopersicon esculentum genomic clone cTOG2002,

```
            genomic survey sequence.
ACCESSION   BH017468
VERSION     BH017468.1  GI:14148506
KEYWORDS    GSS.
SOURCE      Lycopersicon esculentum (Solanum lycopersicum)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 460)
  AUTHORS   van der Hoeven,R., Sun,H., Cho,J., Utterback,T., Ronning,C. and
            Tanksley,S.
  TITLE     Tomato Demethylated Genomic DNA Sequences
  JOURNAL   Unpublished (2001)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html
            tomato demethylated genomic DNA
            Seq primer: M13F-R
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .460
                     /organism="Lycopersicon esculentum"
                     /mol_type="genomic DNA"
                     /cultivar="E6203"
                     /db_xref="taxon:4081"
                     /clone="cTOG20O2"
                     /tissue_type="young leaves"
                     /dev_stage="12-14 weeks post harvest"
                     /lab_host="E.coli JM109"
                     /clone_lib="cTOG"
                     /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                     XhoI; This library was made from short EcoRI digested
                     fragments of the genome of Lycopersicon esculentum ligated
                     into pBS (SK-). The fragments were  cloned into the
                     methylation restrictive E.coli strain JM109 with the
                     purpose of enriching the library for non-methylated DNA
                     fragments.  This procedure may enrich the pool of cloned
                     fragments in JM109 cells for sequences representing
                     expressed genes. Average insert size 1.27 kb."
ORIGIN

Alignment Scores:
Pred. No.:              2.65e+03      Length:         460
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     9             Gaps:           0

US-09-858-332G-28 (1-6) x BH017468 (1-460)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        236 CACAATCACAATCACAAT 253

RESULT 412
BQ853517/c
LOCUS       BQ853517                461 bp    mRNA    linear   EST 14-AUG-2002
DEFINITION  QGB20L16.yg.ab1 QG_ABCDI lettuce salinas Lactuca sativa cDNA clone
            QGB20L16, mRNA sequence.
ACCESSION   BQ853517
VERSION     BQ853517.1  GI:22238982
KEYWORDS    EST.
SOURCE      Lactuca sativa
  ORGANISM  Lactuca sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
            Cichorieae; Lactuca.
REFERENCE   1  (bases 1 to 461)
  AUTHORS   Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J.,
            Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y.,
            Lai,Z., Church,S., Jackson,L. and Bradford,K.
  TITLE     Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
            belongs to contig QG_CA_Contig7255, see http://cgpdb.ucdavis.edu/
            for details.
            Plate: QGB20  row: L  column: 16.
FEATURES             Location/Qualifiers
     source          1. .461
                     /organism="Lactuca sativa"
                     /mol_type="mRNA"
                     /cultivar="Salinas"
                     /db_xref="taxon:4236"
                     /clone="QGB20L16"
                     /lab_host="E.coli"
                     /clone_lib="QG_ABCDI lettuce salinas"
                     /note="Vector: pBRcDNASfiAB; The library was constructed
                     from 10 different sources of RNA from a single genotype.
                     Separate cDNAs were generated using primers that
                     incorporated unique 5' and 3' tags to distinguish each
                     source of RNA. cDNAs were then pooled, size-fractionated,
                     directionally cloned into a custom medium-copy vector and
                     transformations made with four size classes to minimize
                     size bias. Details of each source of RNA and library
                     construction can be obtained at http://cgpdb.ucdavis.edu/
                     TAG_TISSUE=flowers post-fertilized
                     TAG_LIB=QG_ABCDI lettuce salinas
                     TAG_SEQ=TGCCATCGGG"
ORIGIN

Alignment Scores:
Pred. No.:              2.65e+03      Length:         461
Score:                  42.00         Matches:        6
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     5             Gaps:           0

US-09-858-332G-28 (1-6) x BQ853517 (1-461)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        236 CACAACCACAACCACAAC 219

RESULT 413
BY249646
LOCUS       BY249646                461 bp    mRNA    linear   EST 10-DEC-2002
DEFINITION  BY249646 RIKEN full-length enriched, visual cortex Mus musculus
            cDNA clone K230326H06 5', mRNA sequence.
ACCESSION   BY249646
VERSION     BY249646.1  GI:26431158
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 461)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
            Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
            Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
            Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
```

Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.

TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL Nature 420, 563-573 (2002)

PUBMED 12466851

COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Tissues were provided by Michela Fagiolini and Takao K. Hensch (Laboratory for Neuronal Circuit Development Brain Science Institute RIKEN 2-1 Hirosawa,Wako-shi,Saitama 351-0198 Japan) whose assistance we gratefully acknowledge. Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

FEATURES Location/Qualifiers
source 1. .461
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="K230326H06"
/tissue_type="visual cortex"
/clone_lib="RIKEN full-length enriched, visual cortex"

ORIGIN

Alignment Scores:

Pred. No.:	2.65e+03	Length:	461
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	5	Gaps:	0

US-09-858-332G-28 (1-6) x BY249646 (1-461)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        338 CACAACCATAACCACAAC 355
```

RESULT 414
BH144416
LOCUS BH144416 461 bp DNA linear GSS 16-AUG-2001
DEFINITION TDGEK76TH cTOG Lycopersicon esculentum genomic clone cTOG29N7, genomic survey sequence.
ACCESSION BH144416
VERSION BH144416.1 GI:15199291
KEYWORDS GSS.
SOURCE Lycopersicon esculentum (Solanum lycopersicum)
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 461)
AUTHORS van der Hoeven,R., Sun,H., Cho,J., Utterback,T., Ronning,C. and Tanksley,S.
TITLE Tomato Demethylated Genomic DNA Sequences
JOURNAL Unpublished (2001)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
tomato demethylated genomic DNA
Insert Length: 1270 Std Error: 0.00
Seq primer: M13F-R
Class: shotgun.

FEATURES Location/Qualifiers
source 1. .461
/organism="Lycopersicon esculentum"
/mol_type="genomic DNA"
/cultivar="E6203"
/db_xref="taxon:4081"
/clone="cTOG29N7"
/tissue_type="young leaves"
/dev_stage="12-14 weeks post harvest"
/lab_host="E.coli JM109"
/clone_lib="cTOG"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; This library was made from short EcoRI digested fragments of the genome of Lycopersicon esculentum ligated into pBS (SK-). The fragments were cloned into the methylation restrictive E.coli strain JM109 with the purpose of enriching the library for non-methylated DNA fragments. This procedure may enrich the pool of cloned fragments in JM109 cells for sequences representing expressed genes. Average insert size 1.27 kb."

ORIGIN

Alignment Scores:

Pred. No.:	2.65e+03	Length:	461
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

US-09-858-332G-28 (1-6) x BH144416 (1-461)

```
Qy          1 HisAsnHisAsnHisAsn 6
```

```
                  ||||||||||||||||||
Db        237 CACAATCACAATCACAAT 254

RESULT 415
BP926853
LOCUS       BP926853                462 bp    mRNA    linear   EST 23-FEB-2005
DEFINITION  BP926853 full-length enriched poplar cDNA library Populus nigra
            cDNA clone PnFL1-064_M07.f 5', mRNA sequence.
ACCESSION   BP926853
VERSION     BP926853.1  GI:60208483
KEYWORDS    EST.
SOURCE      Populus nigra
  ORGANISM  Populus nigra
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Malpighiales; Salicaceae; Saliceae; Populus.
REFERENCE   1  (bases 1 to 462)
  AUTHORS   Nanjo,T., Futamura,N., Nishiguchi,M., Igasaki,T., Shinozaki,K. and
            Shinohara,K.
  TITLE     Characterization of full-length enriched expressed sequence tags of
            stress-treated poplar leaves
  JOURNAL   Plant Cell Physiol. 45 (12), 1738-1748 (2004)
   PUBMED   15653793
COMMENT     Contact: Tokihiko Nanjo
            Molecular and Cell Biology
            Forestry and Forest Products Research Institute (FFPRI)
            1 Matsunosato, Tsukuba, Ibaraki, 305-8687, Japan
            Tel: 81-29-873-3211
            Fax: 81-29-873-0507
            Email: nanjo@affrc.go.jp.
FEATURES             Location/Qualifiers
     source          1. .462
                     /organism="Populus nigra"
                     /mol_type="mRNA"
                     /db_xref="taxon:3691"
                     /clone="PnFL1-064_M07.f"
                     /sex="female"
                     /tissue_type="leaf"
                     /dev_stage="juvenile"
                     /clone_lib="full-length enriched poplar cDNA library"
                     /note="synonym: Populus nigra var. italica"
ORIGIN

Alignment Scores:
Pred. No.:              2.66e+03        Length:         462
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-858-332G-28 (1-6) x BP926853 (1-462)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         76 CACAATCACAACCACAAC 93

RESULT 416
BM931692/c
LOCUS       BM931692                463 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  UI-E-EJ1-ajg-o-24-0-UI.r1 UI-E-EJ1 Homo sapiens cDNA clone
            UI-E-EJ1-ajg-o-24-0-UI 5', mRNA sequence.
ACCESSION   BM931692
VERSION     BM931692.1  GI:19390865
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 463)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
   PUBMED   8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road , 4156  MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            Tissue Procurement: Dr. Gregg Hageman
             cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa
             Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
             The following repetitive elements were found in this cDNA
            sequence: 434-459, >POLY_A#Simple_repeat
            Seq primer: M13 REVERSE.
FEATURES             Location/Qualifiers
     source          1. .463
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-E-EJ1-ajg-o-24-0-UI"
                     /tissue_type="fetal eyes, lens, eye anterior segment,
                     optic nerve, retina, Retina Foveal and Macular, RPE and
                     Choroid"
                     /dev_stage="fetal and adult"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /clone_lib="UI-E-EJ1"
                     /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-E-EJ1 is a subtracted cDNA library constructed
                     according to Bonaldo, Lennon and Soares, Genome Research,
                     6:791-806, 1996. First strand cDNA synthesis was primed
                     with an oligo-dT primer containing a Not I site. Double
                     stranded cDNA was ligated to an EcoR I adaptor, digested
                     with Not I, and cloned directionally into pT7T3-Pac
                     vector. The oligonucleotide used to prime the synthesis of
                     first-strand cDNA contains a library tag sequence that is
                     located between the Not I site and the (dT)18 tail. The
                     sequence tags for this library are: fetal eyes,
                     AGAATCAAGA; lens, CGATTAGCGA; eye anterior segment,
                     AATGCCGCAT; optic nerve, CCATTAAGTG; retina, CCGCG; Retina
                     Foveal and Macular, GTCC; RPE and Choroid, ACCTA. This
                     library was created for the program, Gene Discovery in the
                     Visual System, supported by National Eye Institute (NEI)."
ORIGIN

Alignment Scores:
Pred. No.:              2.67e+03        Length:         463
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-858-332G-28 (1-6) x BM931692 (1-463)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        243 CATAATCATAATCATAAT 226

RESULT 417
BP043338/c
LOCUS       BP043338                463 bp    mRNA    linear   EST 19-AUG-2004
DEFINITION  BP043338 Lotus corniculatus var. japonicus flower bud Lotus
            corniculatus var. japonicus cDNA clone MFBL041e09_f 3', mRNA
            sequence.
ACCESSION   BP043338
VERSION     BP043338.1  GI:45577118
KEYWORDS    EST.
```

```
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1  (bases 1 to 463)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Characteristics of the Lotus japonicus gene repertoire deduced from
            large-scale expressed sequence tag (EST) analysis
  JOURNAL   Plant Mol. Biol. 54 (3), 405-414 (2004)
   PUBMED   15284495
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .463
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="mRNA"
                     /isolate="Miyakojima MG-20"
                     /db_xref="taxon:34305"
                     /clone="MFBL041e09_f"
                     /tissue_type="flower bud"
                     /clone_lib="Lotus corniculatus var. japonicus flower bud"
ORIGIN

Alignment Scores:
Pred. No.:               2.67e+03        Length:          463
Score:                   42.00           Matches:         6
Percent Similarity:      100.00%         Conservative:    0
Best Local Similarity:   100.00%         Mismatches:      0
Query Match:             100.00%         Indels:          0
DB:                      3               Gaps:            0

US-09-858-332G-28 (1-6) x BP043338 (1-463)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         73 CACAACCACAACCACAAC 56

RESULT 418
CF992233
LOCUS       CF992233                463 bp    mRNA    linear   EST 04-FEB-2005
DEFINITION  16866rsicee_14487.y1 Oryza sativa cv. LYP9 tillering whole plant
            cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF992233
VERSION     CF992233.1  GI:58593925
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 463)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 463
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .463
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="tillering"
                     /clone_lib="Oryza sativa cv. LYP9 tillering whole plant
                     cDNA library"
ORIGIN

Alignment Scores:
Pred. No.:               2.67e+03        Length:          463
Score:                   42.00           Matches:         6
Percent Similarity:      100.00%         Conservative:    0
Best Local Similarity:   100.00%         Mismatches:      0
Query Match:             100.00%         Indels:          0
DB:                      7               Gaps:            0

US-09-858-332G-28 (1-6) x CF992233 (1-463)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        216 CACAACCACAACCACAAC 233

RESULT 419
BF293786
LOCUS       BF293786                464 bp    mRNA    linear   EST 17-NOV-2000
DEFINITION  WHE2153_A03_B05ZS Triticum turgidum L. var. durum (durum wheat)
            whole plant cDNA library Triticum turgidum cDNA clone
            WHE2153_A03_B05, mRNA sequence.
ACCESSION   BF293786
VERSION     BF293786.1  GI:11224850
KEYWORDS    EST.
SOURCE      Triticum turgidum
  ORGANISM  Triticum turgidum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 464)
  AUTHORS   Akhunov,E., Anderson,O.D., Chao,S., Chin,A., Choi,D.W., Close,T.J.,
            Fenton,R.D., Han,P.S., Hsia,C.C., Kang,Y., Kianian,P., Lazo,G.R.,
            Miller,R., Otto,C., Rausch,C.J., Seaton,C.L., Simons,K., Tong,J.C.
            and Zhang,D.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Whole plant cDNA library from Triticum turgidum L. var.
            durum
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene SK primer.
FEATURES             Location/Qualifiers
     source          1. .464
                     /organism="Triticum turgidum"
                     /mol_type="mRNA"
                     /cultivar="Langdon-16"
                     /db_xref="taxon:4571"
                     /clone="WHE2153_A03_B05"
                     /tissue_type="All tissues"
                     /dev_stage="Different growth stages"
                     /lab_host="E. coli SOLR"
```

/clone_lib="Triticum turgidum L. var. durum (durum wheat) whole plant cDNA library"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site_1: EcoRI; Site_2: XhoI; Plants were grown in a growth chamber at North Dakota State University (Kianian, Otto, Simons). Tissues collected from seven-day etiolated seedling leaf, stem, root and seed; leaf from plant at fourth leaf stage; spike from pre-anthesis through 20 days after anthesis; flag leaf; leaf and stem tissue from tillers. and root. Total RNA and poly(A) RNA were prepared from each tissue and then pooled, a cDNA library was made, and the cDNA clones were in vivo excised to give pBluescript phagemids in the TJ Close lab (Akhunov, Chin, Choi, Close, Fenton, Kianian, Otto, Simons, Zhang) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

ORIGIN

Alignment Scores:

Pred. No.:	2.67e+03	Length:	464
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-28 (1-6) x BF293786 (1-464)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        307 CACAACCACAACCACAAT 324
```

RESULT 420
BG088461
LOCUS BG088461 464 bp mRNA linear EST 18-DEC-2003
DEFINITION H3153D07-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone H3153D07 5', mRNA sequence.
ACCESSION BG088461
VERSION BG088461.2 GI:40113128
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 464)
AUTHORS Tanaka,T.S., Jaradat,S.A., Lim,M.K., Kargul,G.J., Wang,X., Grahovac,M.J., Pantano,S., Sano,Y., Piao,Y., Nagaraja,R., Doi,H., Wood,W.H. III, Becker,K.G. and Ko,M.S.H.
TITLE Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)
PUBMED 10922068
COMMENT On Jan 26, 2001 this sequence version replaced gi:12571025.
Other_ESTs: H3153D07-3
Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@lgsun.grc.nia.nih.gov
This clone set has been freely distributed to the community. Please visit http://lgsun.grc.nia.nih.gov/cDNA/15k.html for details.
Plate: H3153 row: D column: 07
Seq primer: -21M13 Reverse
High quality sequence stop: 464
POLYA=No.
FEATURES Location/Qualifiers
source 1. .464
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="niaEST:H3153D07-5"
/db_xref="taxon:10090"
/clone="H3153D07"
/sex="Clones arrayed from a variety of cDNA libraries"
/dev_stage="Clones arrayed from a variety of cDNA libraries"
/lab_host="DH10B"
/clone_lib="NIA Mouse 15K cDNA Clone Set"
/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This clone is among a rearrayed set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos, and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dT)-Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse develolpment, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."

ORIGIN

Alignment Scores:

Pred. No.:	2.67e+03	Length:	464
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-28 (1-6) x BG088461 (1-464)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        157 CACAACCATAACCACAAC 174
```

RESULT 421
BG140455
LOCUS BG140455 464 bp mRNA linear EST 31-JAN-2001
DEFINITION EST480897 wild tomato pollen Lycopersicon pennellii cDNA clone cLPP17C18 5' sequence, mRNA sequence.
ACCESSION BG140455
VERSION BG140455.1 GI:12640644
KEYWORDS EST.
SOURCE Lycopersicon pennellii (Solanum pennellii)
ORGANISM Lycopersicon pennellii
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 464)
AUTHORS van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Utterback,T., Hansen,C., Ronning,C. and Tanksley,S.
TITLE Generation of ESTs from wild tomato (L. pennellii) pollen
JOURNAL Unpublished (2001)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html.
FEATURES Location/Qualifiers
source 1. .464
/organism="Lycopersicon pennellii"
/mol_type="mRNA"
/cultivar="TA56"
/db_xref="taxon:28526"
/clone="cLPP17C18"
/tissue_type="pollen"

/dev_stage="pollen collected from open flowers"
/lab_host="SOLR"
/clone_lib="wild tomato pollen"
/note="Vector: pBluescript SK(-); Site_1: EcoR1; Site_2: Xho1; Pollen was collected from open flowers from L.pennellii TA56, and stored at -80 C until library construction."

ORIGIN

Alignment Scores:

Pred. No.:	2.67e+03	Length:	464
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-28 (1-6) x BG140455 (1-464)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        377 CACAACCACAACCACAAC 394
```

RESULT 422
BQ589046
LOCUS BQ589046 464 bp mRNA linear EST 06-DEC-2002
DEFINITION E012581-024-016-A06-SP6 MPIZ-ADIS-024-storage root Beta vulgaris cDNA clone 024-016-A06 5-PRIME, mRNA sequence.
ACCESSION BQ589046
VERSION BQ589046.1 GI:26118629
KEYWORDS EST.
SOURCE Beta vulgaris
ORGANISM Beta vulgaris
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; Caryophyllales; Amaranthaceae; Beta.
REFERENCE 1 (bases 1 to 464)
AUTHORS Herwig,R., Schulz,B., Weisshaar,B., Hennig,S., Steinfath,M., Drungowski,M., Stahl,D., Wruck,W., Menze,A., O'Brien,J., Lehrach,H. and Radelof,U.
TITLE Construction of a 'unigene' cDNA clone set by oligonucleotide fingerprinting allows access to 25 000 potential sugar beet genes
JOURNAL Plant J. 32 (5), 845-857 (2002)
PUBMED 12472698
COMMENT Contact: Weisshaar B
ADIS DNA core facility at MPIZ
Max-Planck-Institute for Plant Breeding Research
Carl-von-Linne Weg 10, 50829 Koeln, Germany
Fax: 00492215062851
Email: weisshaa@mpiz-koeln.mpg.de
Insert Length: 464 Std Error: 0.00
Plate: 16 row: A column: 06
Seq primer: SP6; CATACGATTTAGGTGACACTATAG.
FEATURES Location/Qualifiers
source 1. .464
/organism="Beta vulgaris"
/mol_type="mRNA"
/cultivar="KWS2320 (double haploid, monogerm breeding line)"
/db_xref="GABI:188019"
/db_xref="taxon:161934"
/clone="024-016-A06"
/tissue_type="storage root"
/lab_host="EMDH10B"
/clone_lib="MPIZ-ADIS-024-storage root"
/note="Vector: pCMVSPORT6; Site_1: SalI; Site_2: NotI; cDNA library from sugar beet, library provided by KWS Kleinwanzlebener Saatzucht AG Einbeck, Germany, contact: b.schulz@kws.de; cloning sites SalI-NotI, primer sites and orientation:
SP6-SalI-CCACGCGTCCG-5prime-cDNA-polyA-CC-NotI-T7; Note: Sequencing granted in the context of the GABI-Beet project, local PI: Dr. Katharina Schneider, coordinator: Prof. Christian Jung; Sequence submission managed by RZPD/GABI-Primary database: http://gabi.rzpd.de"

ORIGIN

Alignment Scores:

Pred. No.:	2.67e+03	Length:	464
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	5	Gaps:	0

US-09-858-332G-28 (1-6) x BQ589046 (1-464)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         80 CATAACCATAACCACAAC 97
```

RESULT 423
BG277124
LOCUS BG277124 465 bp mRNA linear EST 21-FEB-2001
DEFINITION ux63g03.y1 Soares_NKWMD_mandible Mus musculus cDNA clone IMAGE:3514949 5' similar to TR:Q9WTZ9 Q9WTZ9 BRAIN EXPRESSED X-LINKED PROTEIN 3. ;, mRNA sequence.
ACCESSION BG277124
VERSION BG277124.1 GI:13072112
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 465)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Other_ESTs: ux63g03.x1
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1395525
Seq primer: -40RP from Gibco
High quality sequence stop: 369.
FEATURES Location/Qualifiers
source 1. .465
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:3514949"
/tissue_type="mandible"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Soares_NKWMD_mandible"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: NotI; Site_2: EcoRI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCCTTAATTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo. "

ORIGIN

Alignment Scores:

Pred. No.:	2.68e+03	Length:	465
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-28 (1-6) x BG277124 (1-465)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        241 CACAACCATAACCACAAC 258
```

RESULT 424
BJ428195/c
LOCUS BJ428195 465 bp mRNA linear EST 11-MAR-2002
DEFINITION BJ428195 Dictyostelium discoideum cDNA library, VF Dictyostelium discoideum cDNA clone ddv11f06 3', mRNA sequence.
ACCESSION BJ428195
VERSION BJ428195.1 GI:19345552
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 465)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the vegetative stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES Location/Qualifiers
source 1. .465
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddv11f06"
/sex="mat A"
/dev_stage="Growth phase"
/clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:

Pred. No.:	2.68e+03	Length:	465
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	3	Gaps:	0

US-09-858-332G-28 (1-6) x BJ428195 (1-465)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        390 CACAATCACAATCACAAT 373
```

RESULT 425
BJ433147/c
LOCUS BJ433147 465 bp mRNA linear EST 13-MAR-2002
DEFINITION BJ433147 Dictyostelium discoideum cDNA library, VF Dictyostelium discoideum cDNA clone ddv20m22 3', mRNA sequence.
ACCESSION BJ433147
VERSION BJ433147.1 GI:19407869
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 465)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the vegetative stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES Location/Qualifiers
source 1. .465
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddv20m22"
/sex="mat A"
/dev_stage="Growth phase"
/clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:

Pred. No.:	2.68e+03	Length:	465
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	3	Gaps:	0

US-09-858-332G-28 (1-6) x BJ433147 (1-465)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        465 CACAATCACAATCACAAT 448
```

RESULT 426
BG154056
LOCUS BG154056 466 bp mRNA linear EST 05-FEB-2001
DEFINITION 342 L1N01 Lupinus luteus cDNA, mRNA sequence.
ACCESSION BG154056
VERSION BG154056.1 GI:12666086
KEYWORDS EST.
SOURCE Lupinus luteus (yellow lupine)
ORGANISM Lupinus luteus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Genisteae; Lupinus.
REFERENCE 1 (bases 1 to 466)
AUTHORS Podkowinski,J., Kisiel,A., Grabowska,B. and Legocki,A.B.
TITLE ESTs from the early nodules library of Lupinus luteus
JOURNAL Unpublished (2000)
COMMENT Contact: Podkowinski J
Laboratory of Plant Molecular Biology
Institute of Bioorganic Chemistry of Polish Academy of Sciences
Noskowskiego 12/14, 61-704 Poznan, Poland
Tel: 0048 61 8528919
Fax: 0048 61 8520532
Email: jantsp@ibch.poznan.pl
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .466
/organism="Lupinus luteus"
/mol_type="mRNA"
/cultivar="ventus"
/db_xref="taxon:3873"
/tissue_type="roots with young nodules"
/dev_stage="roots with immature nodules harvested seven days post inoculation with Bradyrhizobium sp. WM9 (Lupinus)"
/lab_host="E. coli strain SOLR"
/clone_lib="L1N01"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA from roots with developing nodules harvested seven days post inoculation with Bradyrhizobium sp. WM9 (Lupinus). The cDNA was directionally ligated into the Uni-ZAP XR vector

from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using Ex-Assist helper phage and propagated in SOLR cells."

ORIGIN

Alignment Scores:

Pred. No.:	2.68e+03	Length:	466
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-28 (1-6) x BG154056 (1-466)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        119 CACAATCACAATCACAAT 136
```

RESULT 427
BQ246970
LOCUS BQ246970 466 bp mRNA linear EST 03-MAY-2002
DEFINITION TaE15002H08R TaE15 Triticum aestivum cDNA clone TaE15002H08R, mRNA sequence.
ACCESSION BQ246970
VERSION BQ246970.1 GI:20442846
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 466)
AUTHORS Cloutier,S.
TITLE Wheat functional genomics - Glenlea developing seeds cDNA libraries
JOURNAL Unpublished (2002)
COMMENT Contact: Dr. Sylvie Cloutier
Cereal Research Centre, Agriculture and Agri-food Canada
195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
Tel: (204) 983-2340
Fax: (204) 983-4604
Email: scloutier@agr.gc.ca
was cloned directionally, not all sequences generated with reverse primer were from the 5' end (same with forward primer and 3' end). Average insert size is >1.4 kb
Plate: 002 row: H column: 08
Seq primer: M13 Reverse.
FEATURES Location/Qualifiers
source 1. .466
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Glenlea"
/db_xref="taxon:4565"
/clone="TaE15002H08R"
/tissue_type="developing seeds"
/dev_stage="15 days after anthesis"
/lab_host="E. coli DH10B"
/clone_lib="TaE15"
/note="Vector: pCMV-SPORT6.0 (Invitrogen Technologies); Site_1: NotI; Site_2: MluI; mRNA obtained from wheat seeds of cultivar Glenlea 15 days post-anthesis"
ORIGIN

Alignment Scores:

Pred. No.:	2.68e+03	Length:	466
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	3	Gaps:	0

US-09-858-332G-28 (1-6) x BQ246970 (1-466)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        329 CACAACCACAACCACAAT 346
```

RESULT 428
BQ982354/c
LOCUS BQ982354 466 bp mRNA linear EST 21-AUG-2002
DEFINITION QGE16D15.yg.ab1 QG_EFGHJ lettuce serriola Lactuca sativa cDNA clone QGE16D15, mRNA sequence.
ACCESSION BQ982354
VERSION BQ982354.1 GI:22399879
KEYWORDS EST.
SOURCE Lactuca sativa
ORGANISM Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; campanulids; Asterales; Asteraceae; Cichorioideae; Cichorieae; Lactuca.
REFERENCE 1 (bases 1 to 466)
AUTHORS Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L., Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J., Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z., Church,S., Jackson,L. and Bradford,K.
TITLE Lettuce and Sunflower ESTs from the Compositae Genome Project http://compgenomics.ucdavis.edu/
JOURNAL Unpublished (2002)
COMMENT Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
belongs to contig QG_CA_Contig1494, see http://cgpdb.ucdavis.edu/ for details.
Plate: QGE16 row: D column: 15.
FEATURES Location/Qualifiers
source 1. .466
/organism="Lactuca sativa"
/mol_type="mRNA"
/cultivar="L.serriola"
/db_xref="taxon:4236"
/clone="QGE16D15"
/lab_host="E.coli"
/clone_lib="QG_EFGHJ lettuce serriola"
/note="Vector: pBRcDNASfiAB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at http://cgpdb.ucdavis.edu/
TAG_TISSUE=chemical induction
TAG_LIB=QG_EFGHJ lettuce serriola
TAG_SEQ=TGTAGCCGGG"
ORIGIN

Alignment Scores:

Pred. No.:	2.68e+03	Length:	466
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	5	Gaps:	0

US-09-858-332G-28 (1-6) x BQ982354 (1-466)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        289 CACAATCACAATCACAAT 272
```

```
RESULT 429
CB714934
LOCUS       CB714934                466 bp    mRNA    linear   EST 10-APR-2003
DEFINITION  AMGNNUC:MRPE3-00062-C4-A placenta embryo D17 (10379) Rattus
            norvegicus cDNA clone mrpe3-00062-c4 5', mRNA sequence.
ACCESSION   CB714934
VERSION     CB714934.1  GI:29772082
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 466)
  AUTHORS   Amgen EST Program.
  TITLE     Amgen Rat EST Program
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Dan Fitzpatrick
            Amgen, Inc
            One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA
            Tel: 805 447-4881
            Plate: 00062  row: c  column: 4.
FEATURES             Location/Qualifiers
     source          1. .466
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
                     /clone="mrpe3-00062-c4"
                     /tissue_type="placenta embryo"
                     /clone_lib="placenta embryo D17 (10379)"
                     /note="Vector: pSPORT1; placenta embryo D17"
ORIGIN

Alignment Scores:
Pred. No.:              2.68e+03        Length:         466
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x CB714934 (1-466)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        405 CACAACCACAACCACAAC 422

RESULT 430
CO344556
LOCUS       CO344556                466 bp    mRNA    linear   EST 29-JUN-2004
DEFINITION  EP19507.3prime Exelixis FlyTag CK02 pCDNA-SK+ Drosophila
            melanogaster cDNA clone EP19507 3, mRNA sequence.
ACCESSION   CO344556
VERSION     CO344556.1  GI:49404863
KEYWORDS    EST.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 466)
  AUTHORS   Platt,D., Kopczynski,C., Muzong,C., Laufer,A., Leung,W.,
            Peterson,E. and Swimmer,C.
  TITLE     Exelixis FlyTag EST Project CK02 Library
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: EP.195  row: A  column: 7
            High quality sequence stop: 293.
FEATURES             Location/Qualifiers
     source          1. .466
                     /organism="Drosophila melanogaster"
                     /mol_type="mRNA"
                     /db_xref="taxon:7227"
                     /clone="EP19507"
                     /clone_lib="Exelixis FlyTag CK02 pCDNA-SK+"
                     /note="Organ: mixed stage embryos, imaginal disks, and
                     adult heads; Vector: pCDNA-SK+; Site_1: NotI; Site_2:
                     XhoI; Random primed, normalized library from mixed stage
                     embryos, imaginal disks, and adult heads. Subset of
                     Exelixis FlyTag CK01 clones sequenced from 3' end"
ORIGIN

Alignment Scores:
Pred. No.:              2.68e+03        Length:         466
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     7               Gaps:           0

US-09-858-332G-28 (1-6) x CO344556 (1-466)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        180 CATAATCATAATCATAAT 197

RESULT 431
CV056029
LOCUS       CV056029                466 bp    mRNA    linear   EST 24-AUG-2004
DEFINITION  BNEL13B4 Barley EST endosperm library Hordeum vulgare subsp.
            vulgare cDNA clone BNEL13B4 5' similar to Unknown Function, mRNA
            sequence.
ACCESSION   CV056029
VERSION     CV056029.1  GI:51519168
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 466)
  AUTHORS   Ali,S, Holloway,B. and Taylor,W.C.
  TITLE     Normalisation of cereal endosperm EST libraries for structural and
            functional genomic analysis
  JOURNAL   Plant Mol. Biol. Rep. 18, 123-132 (2000)
COMMENT     Contact: Bill Taylor
            Commonwealth Scientific and Industrial Research Organisation
            Division of Plant Industry.
            CSIRO Plant Industry, GPO Box 1600, Canberra, ACT 2601, Australia
            Tel: 61 2 6246 5223
            Fax: 61 2 6246 5000
            Email: Bill.Taylor@csiro.au
            Seq primer: M13 reverse primer
            High quality sequence stop: 466.
FEATURES             Location/Qualifiers
     source          1. .466
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Himalaya"
                     /sub_species="vulgare"
                     /db_xref="taxon:112509"
                     /clone="BNEL13B4"
                     /tissue_type="endosperm"
                     /dev_stage="developing endosperm tissue 10, 12, 15 dpa
                     (days post anthesis)"
                     /lab_host="DH10B (Life Technology)"
                     /clone_lib="Barley EST endosperm library"
                     /note="Vector: _ZipLox; Site_1: Sal I; Site_2: Not I; mRNA
                     was prepared from endosperm tissues of the barley cultivar
                     Himalaya. cDNA was synthesised from pooled 10, 12, and 15
```

dpa endosperm using Not I-oligo(dT)18 primer/adapter (Pharmacia Biotech), and then ligated to the Sal I-Not I site of _ZipLox vector (Life Technology) after adding a Sal I-Xho I adapter (Stratagene). Constructed by Shahjahan Ali and Bill Taylor."

ORIGIN

Alignment Scores:

Pred. No.:	2.68e+03	Length:	466
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	7	Gaps:	0

US-09-858-332G-28 (1-6) x CV056029 (1-466)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        200 CACAACCACAACCACAAC 217
```

RESULT 432
CV057634
LOCUS CV057634 466 bp mRNA linear EST 24-AUG-2004
DEFINITION BNEL2A7 Barley EST endosperm library Hordeum vulgare subsp. vulgare cDNA clone BNEL2A7 5' similar to Unknown Function, mRNA sequence.
ACCESSION CV057634
VERSION CV057634.1 GI:51520773
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 466)
AUTHORS Ali,S, Holloway,B. and Taylor,W.C.
TITLE Normalisation of cereal endosperm EST libraries for structural and functional genomic analysis
JOURNAL Plant Mol. Biol. Rep. 18, 123-132 (2000)
COMMENT Contact: Bill Taylor
Commonwealth Scientific and Industrial Research Organisation
Division of Plant Industry.
CSIRO Plant Industry, GPO Box 1600, Canberra, ACT 2601, Australia
Tel: 61 2 6246 5223
Fax: 61 2 6246 5000
Email: Bill.Taylor@csiro.au
Seq primer: M13 reverse primer
High quality sequence stop: 466.
FEATURES Location/Qualifiers
source 1. .466
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Himalaya"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="BNEL2A7"
/tissue_type="endosperm"
/dev_stage="developing endosperm tissue 10, 12, 15 dpa (days post anthesis)"
/lab_host="DH10B (Life Technology)"
/clone_lib="Barley EST endosperm library"
/note="Vector: _ZipLox; Site_1: Sal I; Site_2: Not I; mRNA was prepared from endosperm tissues of the barley cultivar Himalaya. cDNA was synthesised from pooled 10, 12, and 15 dpa endosperm using Not I-oligo(dT)18 primer/adapter (Pharmacia Biotech), and then ligated to the Sal I-Not I site of _ZipLox vector (Life Technology) after adding a Sal I-Xho I adapter (Stratagene). Constructed by Shahjahan Ali and Bill Taylor."
ORIGIN

Alignment Scores:

Pred. No.:	2.68e+03	Length:	466
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	7	Gaps:	0

US-09-858-332G-28 (1-6) x CV057634 (1-466)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        200 CACAACCACAACCACAAC 217
```

RESULT 433
CV060702
LOCUS CV060702 466 bp mRNA linear EST 24-AUG-2004
DEFINITION BNEL60b9 Barley EST endosperm library Hordeum vulgare subsp. vulgare cDNA clone BNEL60b9 5' similar to Unknown Function, mRNA sequence.
ACCESSION CV060702
VERSION CV060702.1 GI:51523841
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 466)
AUTHORS Ali,S, Holloway,B. and Taylor,W.C.
TITLE Normalisation of cereal endosperm EST libraries for structural and functional genomic analysis
JOURNAL Plant Mol. Biol. Rep. 18, 123-132 (2000)
COMMENT Contact: Bill Taylor
Commonwealth Scientific and Industrial Research Organisation
Division of Plant Industry.
CSIRO Plant Industry, GPO Box 1600, Canberra, ACT 2601, Australia
Tel: 61 2 6246 5223
Fax: 61 2 6246 5000
Email: Bill.Taylor@csiro.au
Seq primer: M13 reverse primer
High quality sequence stop: 466.
FEATURES Location/Qualifiers
source 1. .466
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Himalaya"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="BNEL60b9"
/tissue_type="endosperm"
/dev_stage="developing endosperm tissue 10, 12, 15 dpa (days post anthesis)"
/lab_host="DH10B (Life Technology)"
/clone_lib="Barley EST endosperm library"
/note="Vector: _ZipLox; Site_1: Sal I; Site_2: Not I; mRNA was prepared from endosperm tissues of the barley cultivar Himalaya. cDNA was synthesised from pooled 10, 12, and 15 dpa endosperm using Not I-oligo(dT)18 primer/adapter (Pharmacia Biotech), and then ligated to the Sal I-Not I site of _ZipLox vector (Life Technology) after adding a Sal I-Xho I adapter (Stratagene). Constructed by Shahjahan Ali and Bill Taylor."
ORIGIN

Alignment Scores:

Pred. No.:	2.68e+03	Length:	466
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	7	Gaps:	0

US-09-858-332G-28 (1-6) x CV060702 (1-466)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        200 CACAACCACAACCACAAC 217

RESULT 434
CV061380
LOCUS       CV061380                466 bp    mRNA    linear   EST 24-AUG-2004
DEFINITION  BNEL68b7 Barley EST endosperm library Hordeum vulgare subsp.
            vulgare cDNA clone BNEL68b7 5' similar to Unknown Function, mRNA
            sequence.
ACCESSION   CV061380
VERSION     CV061380.1  GI:51524519
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 466)
  AUTHORS   Ali,S, Holloway,B. and Taylor,W.C.
  TITLE     Normalisation of cereal endosperm EST libraries for structural and
            functional genomic analysis
  JOURNAL   Plant Mol. Biol. Rep. 18, 123-132 (2000)
COMMENT     Contact: Bill Taylor
            Commonwealth Scientific and Industrial Research Organisation
            Division of Plant Industry.
            CSIRO Plant Industry, GPO Box 1600, Canberra, ACT 2601, Australia
            Tel: 61 2 6246 5223
            Fax: 61 2 6246 5000
            Email: Bill.Taylor@csiro.au
            Seq primer: M13 reverse primer
            High quality sequence stop: 466.
FEATURES             Location/Qualifiers
     source          1. .466
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Himalaya"
                     /sub_species="vulgare"
                     /db_xref="taxon:112509"
                     /clone="BNEL68b7"
                     /tissue_type="endosperm"
                     /dev_stage="developing endosperm tissue 10, 12, 15 dpa
                     (days post anthesis)"
                     /lab_host="DH10B (Life Technology)"
                     /clone_lib="Barley EST endosperm library"
                     /note="Vector: _ZipLox; Site_1: Sal I; Site_2: Not I; mRNA
                     was prepared from endosperm tissues of the barley cultivar
                     Himalaya. cDNA was synthesised from pooled 10, 12, and 15
                     dpa endosperm using Not I-oligo(dT)18 primer/adapter
                     (Pharmacia Biotech), and then ligated to the Sal I-Not I
                     site of _ZipLox vector (Life Technology) after adding a
                     Sal I-Xho I adapter (Stratagene). Constructed by Shahjahan
                     Ali and Bill Taylor."
ORIGIN

Alignment Scores:
Pred. No.:              2.68e+03        Length:         466
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     7               Gaps:           0

US-09-858-332G-28 (1-6) x CV061380 (1-466)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        200 CACAACCACAACCACAAC 217

RESULT 435
W64428
LOCUS       W64428                  466 bp    mRNA    linear   EST 10-JUN-1996
DEFINITION  me07a07.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:386772 5' similar to gb:M38188 OVARIAN GRANULOSA CELL
            13.0 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION   W64428
VERSION     W64428.1  GI:1372079
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 466)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:238604
            Seq primer: mob.REGA+ET
            High quality sequence stop: 322.
FEATURES             Location/Qualifiers
     source          1. .466
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:386772"
                     /sex="unknown"
                     /tissue_type="embryo"
                     /dev_stage="13.5-14.5dpc total fetus"
                     /lab_host="DH10B"
                     /clone_lib="Soares mouse embryo NbME13.5 14.5"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTTT
                     T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                     14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                     State Univ., from 2 ]; double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector. Library went through one round of
                     normalization, and was constructed by Bento Soares and
                     M.Fatima Bonaldo.  "
ORIGIN

Alignment Scores:
Pred. No.:              2.68e+03        Length:         466
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     8               Gaps:           0

US-09-858-332G-28 (1-6) x W64428 (1-466)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        282 CACAACCATAACCACAAC 299

RESULT 436
BI273062
LOCUS       BI273062                467 bp    mRNA    linear   EST 18-JUL-2001
```

```
DEFINITION  NF099B06FL1F1048 Developing flower Medicago truncatula cDNA clone
            NF099B06FL 5', mRNA sequence.
ACCESSION   BI273062
VERSION     BI273062.1  GI:14882924
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 467)
  AUTHORS   Torres-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,
            Flores,H.R., Inman,J.T., Weller,J.W. and May,G.D.
  TITLE     Expressed Sequence Tags from the Samuel Roberts Noble Foundation
            Medicago truncatula flower library
  JOURNAL   Unpublished (2001)
COMMENT     Contact: May GD
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA
            Tel: 580 224 6650
            Fax: 580 224 6692
            Email: gdmay@noble.org
            Insert Length: 467   Std Error: 0.00
            Plate: 099  row: B  column: 06
            Seq primer: TCACACAGGAAACAGCTATGAC.
FEATURES             Location/Qualifiers
     source          1. .467
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /db_xref="taxon:3880"
                     /clone="NF099B06FL"
                     /tissue_type="Developing flowers"
                     /dev_stage="Developmentally pooled. Contains a mixture of
                     very young, developing, fully-opened flowers and flowers
                     in early transition into pods."
                     /clone_lib="Developing flower"
                     /note="Vector: Lambda Zap; cDNA was prepared from polyA+
                     enriched, pooled samples of equivalent  amounts of total
                     RNA from very young, developing, fully-opened flowers  and
                     flowers transitioning into pods. The cDNA was
                     directionally ligated  into the Uni-Zap XR vector
                     (Stratagene) and packaged using the Gigapack  III Gold
                     packaging extracts. Phagemids containing cDNA inserts were
                     in  vivo excised from the recombinant Uni-ZAP XR vector
                     using ExAssist helper  phage and the E. coli strain
                     XL1-Blue MRF' (Stratagene). Excised  plasmids were plated
                     using SOLR cells."
ORIGIN

Alignment Scores:
Pred. No.:                2.69e+03        Length:          467
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       2               Gaps:            0

US-09-858-332G-28 (1-6) x BI273062 (1-467)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        126 CACAACCACAACCACAAC 143

RESULT 437
BB856288
LOCUS       BB856288                 467 bp    mRNA    linear   EST 26-NOV-2001
DEFINITION  BB856288 RIKEN full-length enriched, B16 F10Y cells Mus musculus
            cDNA clone G370033M05 5', mRNA sequence.
ACCESSION   BB856288
VERSION     BB856288.1  GI:17097742
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 467)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K.,
            Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
            Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,
            Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
            2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
            and Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
            Sugahara,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .467
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="G370033M05"
                     /cell_type="B16 F10Y cells"
                     /clone_lib="RIKEN full-length enriched, B16 F10Y cells"
ORIGIN

Alignment Scores:
Pred. No.:                2.69e+03        Length:          467
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       2               Gaps:            0

US-09-858-332G-28 (1-6) x BB856288 (1-467)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        318 CACAACCATAACCACAAC 335

RESULT 438
BJ431306/c
```

```
LOCUS       BJ431306                467 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ431306 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv13o13 3', mRNA sequence.
ACCESSION   BJ431306
VERSION     BJ431306.1  GI:19406028
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 467)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .467
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv13o13"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:              2.69e+03        Length:         467
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-858-332G-28 (1-6) x BJ431306 (1-467)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        211 CACAATCACAATCACAAT 194

RESULT 439
BU545370/c
LOCUS       BU545370                467 bp    mRNA    linear   EST 16-SEP-2002
DEFINITION  GM880004B10F03 Gm-r1088 Glycine max cDNA clone Gm-r1088-1421 3',
            mRNA sequence.
ACCESSION   BU545370
VERSION     BU545370.1  GI:22928231
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 467)
  AUTHORS   Vodkin,L., Shoemaker,R., Keim,P., Retzel,E., Khanna,A., Shealy,R.,
            Clough,S., Thibaud-Nissen,F., Coryell,V., Erpelding,J., Raph,C.,
            Shoop,E., Stromvik,M., Schweitzer,P., Gong,G. and Liu,L.
  TITLE     A Functional Genomics Program for Soybean (NSF 9872565) (2002)
  JOURNAL   Unpublished (2002)
COMMENT     Other_ESTs: AW185434 corresponding to Gm-c1027-147 (5')
            Contact: Vodkin, L.O., PI, A Functional Genomics Program for
            Soybean (NSF 9872565)
            Lewin, H. A., Director, Keck Center for Comparative and Functional
            Genomics
            University of Illinois
            Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
            Tel: (217) 244-6147
            Fax: (217) 333-4582
            Email: l-vodkin@uiuc.edu
            Insert Length: 467   Std Error: 0.00
            Plate: GM880004B10  row: F  column: 03
            Seq primer: 5'-TTTTTTTTTTTTTTTTTTT(A/C/G)-3'
            High quality sequence stop: 467.
FEATURES             Location/Qualifiers
     source          1. .467
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /db_xref="taxon:3847"
                     /clone="Gm-r1088-1421"
                     /clone_lib="Gm-r1088"
                     /note="The library Gm-r1088 is a sequence-driven, reracked
                     set of 9,216 low redundancy clones selected from cDNA
                     libraries from various tissues and stages of development
                     of soybean that consists of 2,706 cDNAs from germinating
                     cotyledons (source library Gm-c1027); 1,355 cDNAS from
                     immature seed coats (libraries Gm-c1019 and Gm-c1023); 917
                     cDNAs from tissue culture derived somatic embryos (source
                     libraries Gm-c1036 and Gm-c1075); 3,275 cDNAs from
                     germinating seedlings, shoot tips, or leaves exposed to
                     various stresses (source libraries Gm-c1064, Gm-c1065,
                     Gm-c1066, and Gm-c1067; and Gm-c1068); and 963 cDNAs from
                     young leaves exposed to bacterial and fungal pathogens
                     (source libraries Gm-c1072, Gm-c1073, and Gm-c1074). The
                     5' ESTs of the source clones from the different libraries
                     was used to select singletons, or a representative of each
                     contig, which were reracked to form library Gm-r1088 and
                     the cDNA clones of the reracked Gm-r1088 library were then
                     sequenced at the 3' end. The unigene selection and 3'
                     sequencing was funded by NSF Plant Genome project #9872565
                     (http://soybeangenomics.cropsci.uiuc.edu/) as part of
                     creation of a low redundancy soybean cDNA set. The source
                     cDNA libraries were constructed by the laboratories of
                     Lila Vodkin, University of Illinois, Randy Shoemaker, Iowa
                     State University, and Paul Keim, Northern Arizona
                     University as part of the Public EST project,
                     http://129.186.26.94/soybeanest.html. The contig analysis
                     to select unique genes was performed by the laboratory of
                     Ernest Retzel, Center for Computational Genomics and
                     Bioinformatics, University of Minnesota,
                     http://web.ahc.umn.edu/biodata/nsfsoy/. Reracking and 3'
                     sequencing were conducted by services of the University of
                     Illinois Keck Center for Comparative and Functional
                     Genomics http://www.biotech.uiuc.edu/keck.htm. Note: The
                     corresponding 5' EST from each clone in the Gm-r1088
                     library is listed in the 'OTHER EST' field. The detailed
                     information on the source library for each clone can also
                     be obtained by referring to the Genome Systems clone ID of
                     the original cDNA library that is also listed under 'OTHER
                     EST'."
ORIGIN

Alignment Scores:
Pred. No.:              2.69e+03        Length:         467
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     5               Gaps:           0

US-09-858-332G-28 (1-6) x BU545370 (1-467)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        215 CATAACCATAACCATAAC 198

RESULT 440
CF974481
LOCUS       CF974481                467 bp    mRNA    linear   EST 24-NOV-2003
```

```
DEFINITION  PSU_np13\G06 Defense-related ESTs from Cacao Leaves Theobroma cacao
            cDNA, mRNA sequence.
ACCESSION   CF974481
VERSION     CF974481.1  GI:38500241
KEYWORDS    EST.
SOURCE      Theobroma cacao (cacao)
  ORGANISM  Theobroma cacao
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
            Theobroma.
REFERENCE   1  (bases 1 to 467)
  AUTHORS   Verica,J.A., Maximova,S.N., Strem,M., Carlson,J., Bailey,B. and
            Guiltinan,M.J.
  TITLE     Isolation of Defense-related ESTs from Subtracted-Normalized cDNA
            Libraries of Elicited Cacao (Theobroma cacao L.) Leaves
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Guiltinan MJ
            Department of Horticulture
            Penn State University
            306 Wartik Lab, Universtiy Park, PA 16801, USA
            Tel: 814 863 7957
            Fax: 814 863 6139
            Email: mjg9@psu.edu.
FEATURES             Location/Qualifiers
     source          1. .467
                     /organism="Theobroma cacao"
                     /mol_type="mRNA"
                     /cultivar="Commun"
                     /db_xref="taxon:3641"
                     /lab_host="E. coli"
                     /clone_lib="Defense-related ESTs from Cacao Leaves"
                     /note="Organ: Leaves; Vector: pCR4-TOPO; Subtracted
                     Normalized EST library from cacao leaves sprayed with the
                     defense elicitors: benzothiadiazole, methyl jasmonate,
                     ethylene, and Nep1"
ORIGIN

Alignment Scores:
Pred. No.:                2.69e+03        Length:          467
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       7               Gaps:            0

US-09-858-332G-28 (1-6) x CF974481 (1-467)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        157 CACAATCACAACCACAAC 174

RESULT 441
CF974699
LOCUS       CF974699                467 bp    mRNA    linear   EST 24-NOV-2003
DEFINITION  PSU_nupup2.2\H09 Defense-related ESTs from Cacao Leaves Theobroma
            cacao cDNA, mRNA sequence.
ACCESSION   CF974699
VERSION     CF974699.1  GI:38500459
KEYWORDS    EST.
SOURCE      Theobroma cacao (cacao)
  ORGANISM  Theobroma cacao
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Malvales; Malvaceae; Byttnerioideae;
            Theobroma.
REFERENCE   1  (bases 1 to 467)
  AUTHORS   Verica,J.A., Maximova,S.N., Strem,M., Carlson,J., Bailey,B. and
            Guiltinan,M.J.
  TITLE     Isolation of Defense-related ESTs from Subtracted-Normalized cDNA
            Libraries of Elicited Cacao (Theobroma cacao L.) Leaves
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Guiltinan MJ
            Department of Horticulture
            Penn State University
            306 Wartik Lab, Universtiy Park, PA 16801, USA
            Tel: 814 863 7957
            Fax: 814 863 6139
            Email: mjg9@psu.edu.
FEATURES             Location/Qualifiers
     source          1. .467
                     /organism="Theobroma cacao"
                     /mol_type="mRNA"
                     /cultivar="Commun"
                     /db_xref="taxon:3641"
                     /lab_host="E. coli"
                     /clone_lib="Defense-related ESTs from Cacao Leaves"
                     /note="Organ: Leaves; Vector: pCR4-TOPO; Subtracted
                     Normalized EST library from cacao leaves sprayed with the
                     defense elicitors: benzothiadiazole, methyl jasmonate,
                     ethylene, and Nep1"
ORIGIN

Alignment Scores:
Pred. No.:                2.69e+03        Length:          467
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       7               Gaps:            0

US-09-858-332G-28 (1-6) x CF974699 (1-467)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        157 CACAATCACAACCACAAC 174

RESULT 442
CV057671
LOCUS       CV057671                467 bp    mRNA    linear   EST 24-AUG-2004
DEFINITION  BNEL2E1 Barley EST endosperm library Hordeum vulgare subsp. vulgare
            cDNA clone BNEL2E1 5' similar to Unknown Function, mRNA sequence.
ACCESSION   CV057671
VERSION     CV057671.1  GI:51520810
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 467)
  AUTHORS   Ali,S, Holloway,B. and Taylor,W.C.
  TITLE     Normalisation of cereal endosperm EST libraries for structural and
            functional genomic analysis
  JOURNAL   Plant Mol. Biol. Rep. 18, 123-132 (2000)
COMMENT     Contact: Bill Taylor
            Commonwealth Scientific and Industrial Research Organisation
            Division of Plant Industry.
            CSIRO Plant Industry, GPO Box 1600, Canberra, ACT 2601, Australia
            Tel: 61 2 6246 5223
            Fax: 61 2 6246 5000
            Email: Bill.Taylor@csiro.au
            Seq primer: M13 reverse primer
            High quality sequence stop: 467.
FEATURES             Location/Qualifiers
     source          1. .467
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Himalaya"
                     /sub_species="vulgare"
                     /db_xref="taxon:112509"
                     /clone="BNEL2E1"
                     /tissue_type="endosperm"
                     /dev_stage="developing endosperm tissue 10, 12, 15 dpa
                     (days post anthesis)"
                     /lab_host="DH10B (Life Technology)"
```

/clone_lib="Barley EST endosperm library"
/note="Vector: _ZipLox; Site_1: Sal I; Site_2: Not I; mRNA was prepared from endosperm tissues of the barley cultivar Himalaya. cDNA was synthesised from pooled 10, 12, and 15 dpa endosperm using Not I-oligo(dT)18 primer/adapter (Pharmacia Biotech), and then ligated to the Sal I-Not I site of _ZipLox vector (Life Technology) after adding a Sal I-Xho I adapter (Stratagene). Constructed by Shahjahan Ali and Bill Taylor."
ORIGIN

Alignment Scores:

Pred. No.:	2.69e+03	Length:	467
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	7	Gaps:	0

US-09-858-332G-28 (1-6) x CV057671 (1-467)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        200 CACAACCACAACCACAAC 217
```

RESULT 443
AQ526634
LOCUS AQ526634 467 bp DNA linear GSS 11-MAY-1999
DEFINITION HS_5193_A1_B07_SP6E RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=769 Col=13 Row=C, genomic survey sequence.
ACCESSION AQ526634
VERSION AQ526634.1 GI:4769784
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 467)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
PUBMED 10449764
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
Plate: 769 row: C column: 13
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 467.
FEATURES Location/Qualifiers
source 1. .467
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=769 Col=13 Row=C"
/sex="male"
/clone_lib="RPCI-11 Human Male BAC Library"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"
ORIGIN

Alignment Scores:

Pred. No.:	2.69e+03	Length:	467
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

US-09-858-332G-28 (1-6) x AQ526634 (1-467)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        291 CATAATCATAATCATAAT 308
```

RESULT 444
AA253897
LOCUS AA253897 468 bp mRNA linear EST 13-MAR-1997
DEFINITION my31e10.r1 Barstead mouse pooled organs MPLRB4 Mus musculus cDNA clone IMAGE:697482 5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION AA253897
VERSION AA253897.1 GI:1888253
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 468)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:431042
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 455.
FEATURES Location/Qualifiers
source 1. .468
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:697482"
/sex="mixed"
/tissue_type="pooled organs"
/dev_stage="7 day"
/lab_host="DH10B"
/clone_lib="Barstead mouse pooled organs MPLRB4"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors [GTTGGATTCGGTACC], digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library constructed by Bob Barstead."

```
ORIGIN

Alignment Scores:
Pred. No.:                2.7e+03        Length:         468
Score:                    42.00          Matches:        6
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       1              Gaps:           0

US-09-858-332G-28 (1-6) x AA253897 (1-468)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        169 CACAACCATAACCACAAC 186

RESULT 445
AW257025
LOCUS       AW257025                 468 bp    mRNA    linear   EST 20-DEC-1999
DEFINITION  EST305162 KV2 Medicago truncatula cDNA clone KV2-6N21, mRNA
            sequence.
ACCESSION   AW257025
VERSION     AW257025.1  GI:6605282
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 468)
  AUTHORS   VandenBosch,K., Hurt,J., Moore,J., Beremand,P., Peng,H., Ellis,L.,
            Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and
            Fraser,C.M.
  TITLE     ESTs from roots of Medicago truncatula after Rhizobium inoculation
  JOURNAL   Unpublished (1999)
COMMENT     Contact: VandenBosch K
            Department of Plant Biology
            University of Minnesota
            220 BioSci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA
            Tel: 612 624 2755
            Fax: 612 625 1738
            Email: kvandenb@cbs.umn.edu
            Texas A&M EST name:T115730e
            TIGR sequence name:MTAAK83TK
            More information is availble at. . (and for clone ordering info)
            http://chrysie.tamu.edu/medicago
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES              Location/Qualifiers
     source           1. .468
                      /organism="Medicago truncatula"
                      /mol_type="mRNA"
                      /cultivar="genotype A17"
                      /db_xref="taxon:3880"
                      /clone="KV2-6N21"
                      /tissue_type="Seedling roots"
                      /dev_stage="2 days post-inoculation with Sinorhizobium
                      meliloti"
                      /lab_host="E. coli strain SOLR"
                      /clone_lib="KV2"
                      /note="Vector: pBluescript SK -; Site_1: EcoRI; Site_2:
                      XhoI; cDNA was prepared from polyA+ enriched RNA.  The
                      cDNA was  directionally ligated into the Unizap XR vector
                      from Stratagene  and packaged using Gigapack III Gold
                      packaging extracts.  Plasmids containing cDNA inserts
                      were excised from the  recombinant lambda-Zap phage using
                      Ex-assist helper phage and  propagated in SOLR cells."
ORIGIN

Alignment Scores:
Pred. No.:                2.7e+03        Length:         468
Score:                    42.00          Matches:        6
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       1              Gaps:           0

US-09-858-332G-28 (1-6) x AW257025 (1-468)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         14 CATAATCACAATCACAAT 31

RESULT 446
BI807757
LOCUS       BI807757                 468 bp    mRNA    linear   EST 02-OCT-2001
DEFINITION  B002D06 Oryza sativa mature leaf library induced by M.grisea Oryza
            sativa cDNA clone B002D06, mRNA sequence.
ACCESSION   BI807757
VERSION     BI807757.1  GI:15854931
KEYWORDS    EST.
SOURCE      Oryza sativa
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 468)
  AUTHORS   Dong,H.T., Li,D.B., Zhuang,X.F., Dai,C.G., Sun,L.X., Pei,Y.X.,
            Wu,H.F., Jiang,Y.X., Yu,F.C., Gao,Q.K. and Lou,Y.C.
  TITLE     A Gene Expression Screen in Oryza sativa
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Haitao Dong, Debao Li
            Bioinfomatics and Gene Network Research Group
            Zhejiang University
            Kaixuan Road 268#, Hangzhou, Zhejiang, P.R.China
            Tel: 0086-571-86892051
            Fax: 0086-571-86961525
            Email: webmaster@estarray.org, URL: http://www.estarray.org
            Seq primer: M13 forward primer.
FEATURES              Location/Qualifiers
     source           1. .468
                      /organism="Oryza sativa"
                      /mol_type="mRNA"
                      /db_xref="taxon:4530"
                      /clone="B002D06"
                      /tissue_type="leaf"
                      /dev_stage="Mature stage"
                      /clone_lib="Oryza sativa mature leaf library induced by
                      M.grisea"
                      /note="Vector: pSport2"
ORIGIN

Alignment Scores:
Pred. No.:                2.7e+03        Length:         468
Score:                    42.00          Matches:        6
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       3              Gaps:           0

US-09-858-332G-28 (1-6) x BI807757 (1-468)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         96 CACAACCACAACCACAAC 113

RESULT 447
CA647588/c
LOCUS       CA647588                 468 bp    mRNA    linear   EST 23-NOV-2002
DEFINITION  wre1n.pk0112.a2 wre1n Triticum aestivum cDNA clone wre1n.pk0112.a2
            5' end, mRNA sequence.
ACCESSION   CA647588
VERSION     CA647588.1  GI:25225884
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
```

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 468)
AUTHORS Tingey,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z., Miao,G., Caraher,N. and Hanafey,M.K.
TITLE DuPont Wheat cDNA Sequence
JOURNAL Unpublished (2002)
COMMENT Contact: Scott V. Tingey
Crop Genetics
E. I. DuPont de Nemours and Company
1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
Tel: 302-631-2602
Fax: 302-631-2607
Email: Scott.V.Tingey@USA.dupont.com
Seq primer: M13.
FEATURES Location/Qualifiers
source 1. .468
/organism="Triticum aestivum"
/mol_type="mRNA"
/db_xref="taxon:4565"
/clone="wre1n.pk0112.a2"
/tissue_type="root"
/clone_lib="wre1n"
/note="Vector: pBluescript SK+; Site_1: EcoRI; Site_2: XhoI; Wheat (Triticum aestivum L.) root; normalized from wre1 library"
ORIGIN

Alignment Scores:

Pred. No.:	2.7e+03	Length:	468
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-09-858-332G-28 (1-6) x CA647588 (1-468)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        242 CATAATCATAATCATAAC 225
```

RESULT 448
CA715672
LOCUS CA715672 468 bp mRNA linear EST 26-NOV-2002
DEFINITION wdk3c.pk008.f18 wdk3c Triticum aestivum cDNA clone wdk3c.pk008.f18 5' end, mRNA sequence.
ACCESSION CA715672
VERSION CA715672.1 GI:25437465
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 468)
AUTHORS Tingey,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z., Miao,G., Caraher,N. and Hanafey,M.K.
TITLE DuPont Wheat cDNA Sequence
JOURNAL Unpublished (2002)
COMMENT Contact: Scott V. Tingey
Crop Genetics
E. I. DuPont de Nemours and Company
1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
Tel: 302-631-2602
Fax: 302-631-2607
Email: Scott.V.Tingey@USA.dupont.com
Seq primer: M13.
FEATURES Location/Qualifiers
source 1. .468
/organism="Triticum aestivum"
/mol_type="mRNA"
/db_xref="taxon:4565"
/clone="wdk3c.pk008.f18"
/tissue_type="kernel"
/lab_host="DH10B"
/clone_lib="wdk3c"
/note="Vector: pBluescript SK+; Site_1: EcoRI; Site_2: XhoI; Wheat (Triticum aestivum L.) developing kernel, 14 days after anthesis."
ORIGIN

Alignment Scores:

Pred. No.:	2.7e+03	Length:	468
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-09-858-332G-28 (1-6) x CA715672 (1-468)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        295 CATAACCATAATCATAAC 312
```

RESULT 449
CX303292
LOCUS CX303292 468 bp mRNA linear EST 06-MAY-2005
DEFINITION C08029F05SK PhyRootSw1 Citrus sinensis cDNA clone C08029F05, mRNA sequence.
ACCESSION CX303292
VERSION CX303292.1 GI:63072146
KEYWORDS EST.
SOURCE Citrus sinensis
ORGANISM Citrus sinensis
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Sapindales; Rutaceae; Citrus.
REFERENCE 1 (bases 1 to 468)
AUTHORS Forment,J., Gadea,J., Huerta,L., Abizanda,L., Agusti,J., Alamar,S., Alos,E., Andres,F., Arribas,R., Beltran,J.P., Berbel,A., Blazquez,M.A., Brumos,J., Canas,L.A., Cercos,M., Colmenero-Flores,J.M., Conesa,A., Estables,B., Gandia,M., Garcia-Martinez,J.L., Gimeno,J., Gisbert,A., Gomez,G., Gonzalez-Candelas,L., Granell,A., Guerri,J., Lafuente,M.T., Madueno,F., Marcos,J.F., Marques,M.C., Martinez,F., Martinez-Godoy,M.A., Miralles,S., Moreno,P., Navarro,L., Pallas,V., Perez-Amador,M.A., Perez-Valle,J., Pons,C., Rodrigo,I., Rodriguez,P.L., Royo,C., Serrano,R., Soler,G., Tadeo,F., Talon,M., Terol,J., Trenor,M., Vaello,L., Vicente,O., Vidal,Ch., Zacarias,L. and Conejero,V.
TITLE Development of a citrus genome-wide EST collection and cDNA microarray as resources for genomic studies
JOURNAL Plant Mol. Biol. 57 (3), 375-391 (2005)
PUBMED 15830128
COMMENT Contact: Forment J
Genomics Laboratory
Instituto de Biologia Molecular y Celular de Plantas (Universidad Politecnica de Valencia - Consejo Superior de Investigaciones Cientificas)
Avenida de los Naranjos s/n, 46022 Valencia, Spain
Email: jforment@ibmcp.upv.es.
FEATURES Location/Qualifiers
source 1. .468
/organism="Citrus sinensis"
/mol_type="mRNA"
/db_xref="taxon:2711"
/clone="C08029F05"
/sex="hermaphrodite"
/dev_stage="2-3 month old seedlings, 20-30 cm. high"
/lab_host="Escherichia coli"
/clone_lib="PhyRootSw1"
/note="Organ: roots; Vector: Uni-ZAP XR; cDNA library made from a mixture of equal amounts of poly-A+ RNA from roots

of plant inoculated with Phytophthora citrophtora zoospores by inmersion in a suspension with 2000 or 15000 zoospores/mL, extracted at 0, 2, 4, 8, 13, and 18 days after inoculation"
ORIGIN

Alignment Scores:

Pred. No.:	2.7e+03	Length:	468
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	8	Gaps:	0

US-09-858-332G-28 (1-6) x CX303292 (1-468)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        223 CACAACCACAACCACAAC 240
```

RESULT 450
CW772270/c
LOCUS CW772270 468 bp DNA linear GSS 12-NOV-2004
DEFINITION OG_BBa0088O11.f OG_BBa Oryza glaberrima genomic clone OG_BBa0088O11 5', genomic survey sequence.
ACCESSION CW772270
VERSION CW772270.1 GI:55711990
KEYWORDS GSS.
SOURCE Oryza glaberrima (African rice)
ORGANISM Oryza glaberrima
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 468)
AUTHORS Kim,H., Yu,Y., Wissotski,M., Byrne,M., Stum,D., Smart,D., Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
TITLE OMAP
JOURNAL Unpublished (2004)
COMMENT Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: rwing@genome.arizona.edu
PCR PRimers
FORWARD: TAA TAC GAC TCA CTA TAG GG
BACKWARD: CAC TCA TTA GGC ACC CCA
Plate: 0088 row: O column: 11
Seq primer: TAA TAC GAC TCA CTA TAG GG
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .468
/organism="Oryza glaberrima"
/mol_type="genomic DNA"
/db_xref="taxon:4538"
/clone="OG_BBa0088O11"
/tissue_type="young leaves"
/lab_host="DH10B T1 phage resistant"
/clone_lib="OG_BBa"
/note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

Alignment Scores:

Pred. No.:	2.7e+03	Length:	468
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	10	Gaps:	0

US-09-858-332G-28 (1-6) x CW772270 (1-468)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        326 CACAACCACAACCACAAC 309
```

RESULT 451
CF370711/c
LOCUS CF370711 469 bp mRNA linear EST 27-AUG-2003
DEFINITION rg52g11.y1 Meloidogyne hapla female SMART pGEM Meloidogyne hapla cDNA 5', mRNA sequence.
ACCESSION CF370711
VERSION CF370711.1 GI:34317957
KEYWORDS EST.
SOURCE Meloidogyne hapla
ORGANISM Meloidogyne hapla
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina; Tylenchoidea; Meloidogynidae; Meloidogyninae; Meloidogyne.
REFERENCE 1 (bases 1 to 469)
AUTHORS McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagareishvili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe,M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE The Washington Univ. Nematode EST Project, 1999
JOURNAL Unpublished (1999)
COMMENT Contact: McCarter JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Cloned unidirectionally. Poly(A)+ RNA was concentrated and purified using Dynabeads (Dynal) and mRNA eluted for first strand synthesis. First strand cDNA was created using MMLV RT (Powerscript, Clontech) and primed with oligo(dT) with XhoI site and 5'SMART 'anchor' added using chimeric DNA-RNA oligo. 12 PCR cycles were done using first strand and primers specific to SMART oligo and 3' end. Double stranded cDNA was digested using XhoI/NotI, fractioned on Chroma-spin 400 columns (Clontech) and ligated to digested pGEM-11zf(+) plasmid. Chemically competent DH10B cells were used as host cells. Library materials provided by Dr. David Bird of North Carolina State University. Library construction by Jeff Rousch. See www.nematode.net for additional project information.
Seq primer: Sp6
High quality sequence stop: 469.
FEATURES Location/Qualifiers
source 1. .469
/organism="Meloidogyne hapla"
/mol_type="mRNA"
/db_xref="taxon:6305"
/tissue_type="whole organism"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="Meloidogyne hapla female SMART pGEM"
/note="Vector: plasmid (ampicillin resistant); Site_1: XhoI; Site_2: NotI; Cloned unidirectionally. Poly(A)+ RNA was concentrated and purified using Dynabeads (Dynal) and mRNA eluted for first strand synthesis. First strand cDNA was created using MMLV RT (Powerscript, Clontech) and primed with oligo(dT) with XhoI site and 5'SMART 'anchor' added using chimeric DNA-RNA oligo. 12 PCR cycles were done using first strand and primers specific to SMART oligo and 3' end. Double stranded cDNA was digested using XhoI/NotI, fractioned on Chroma-spin 400 columns (Clontech) and ligated to digested pGEM-11zf(+) plasmid. Chemically competent DH10B cells were used as host cells. Library materials provided by Dr. David Bird of North Carolina State University. Library construction by Jeff Rousch. See www.nematode.net for additional project information."

```
ORIGIN

Alignment Scores:
Pred. No.:              2.7e+03        Length:         469
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     6              Gaps:           0

US-09-858-332G-28 (1-6) x CF370711 (1-469)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        204 CATAATCACAATCATAAT 187

RESULT 452
CN729649/c
LOCUS       CN729649                469 bp    mRNA    linear   EST 14-MAY-2005
DEFINITION  EX20LIB3_UP_001_B08_16DEC2003_062 Brassica napus EX20LIB3 Brassica
            napus cDNA 5', mRNA sequence.
ACCESSION   CN729649
VERSION     CN729649.1  GI:65287453
KEYWORDS    EST.
SOURCE      Brassica napus (rape)
  ORGANISM  Brassica napus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 469)
  AUTHORS   Georges,F.
  TITLE     Gene expression at early stages of B. napus seed development as
            revealed by transcript profiling under normal and abiotic stress
            conditions
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Fawzy Georges
            Plant Biotechnology Institute, National Research Council Canada
            110 Gymnasium Place, Saskatoon, Saskatchewan, Canada S7N 7R9
            Tel: 306 975 4839
            Fax: 306 975 4815
            Email: Pawzy.Georges@nrc-cnrc.gc.ca
            High quality sequence stop: 469.
FEATURES             Location/Qualifiers
     source          1. .469
                     /organism="Brassica napus"
                     /mol_type="mRNA"
                     /db_xref="taxon:3708"
                     /clone_lib="Brassica napus EX20LIB3"
                     /note="Vector: pGemT-Easy; Sequences have been trimmed to
                     remove vector and low quality regions using LUCY sequence
                     cleanup software (www.tigr.org)."
ORIGIN

Alignment Scores:
Pred. No.:              2.7e+03        Length:         469
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     7              Gaps:           0

US-09-858-332G-28 (1-6) x CN729649 (1-469)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         33 CATAACCATAACCATAAC 16

RESULT 453
BE503307
LOCUS       BE503307                470 bp    mRNA    linear   EST 04-AUG-2000
DEFINITION  7a15e04.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:3218814 3',
            mRNA sequence.
ACCESSION   BE503307
VERSION     BE503307.1  GI:9705715
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 470)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer: -40UP from Gibco
            High quality sequence stop: 462.
FEATURES             Location/Qualifiers
     source          1. .470
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3218814"
                     /tissue_type="pooled germ cell tumors"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_GC6"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA
                     from the normalized library NCI_CGAP_GC4 was prepared, and
                     ss circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
                     of 5,000 clones made from the same library (cloneIDs
                     1257096-1258631, 1469064-1470983, and 1475592-1476743).
                     Subtraction by Bento Soares and M. Fatima Bonaldo.  "
ORIGIN

Alignment Scores:
Pred. No.:              2.71e+03       Length:         470
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     2              Gaps:           0

US-09-858-332G-28 (1-6) x BE503307 (1-470)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        192 CATAATCATAATCATAAT 209

RESULT 454
AV422361/c
LOCUS       AV422361                471 bp    mRNA    linear   EST 10-MAR-2004
DEFINITION  AV422361 Lotus japonicus young plants (two-week old) Lotus
            corniculatus var. japonicus cDNA clone MWM012g01_r 5', mRNA
            sequence.
ACCESSION   AV422361
VERSION     AV422361.1  GI:7777139
KEYWORDS    EST.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
```

```
            Lotus.
REFERENCE   1  (bases 1 to 471)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Generation of 7137 non-redundant expressed sequence tags from a
            legume, Lotus japonicus
  JOURNAL   DNA Res. 7 (2), 127-130 (2000)
   PUBMED   10819328
COMMENT     Contact: Yasukazu Nakamura
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: ynakamu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .471
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="mRNA"
                     /isolate="Miyakojima MG-20"
                     /db_xref="taxon:34305"
                     /clone="MWM012g01_r"
                     /tissue_type="whole plant"
                     /dev_stage="two-week old plant"
                     /clone_lib="Lotus japonicus young plants (two-week old)"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI; synonym: Lotus japonicus"
ORIGIN

Alignment Scores:
Pred. No.:                2.72e+03        Length:         471
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       1               Gaps:           0

US-09-858-332G-28 (1-6) x AV422361 (1-471)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        466 CACAATCACAATCACAAT 449

RESULT 455
BJ380110/c
LOCUS       BJ380110                471 bp    mRNA    linear   EST 08-MAR-2002
DEFINITION  BJ380110 Dictyostelium discoideum cDNA library, CF Dictyostelium
            discoideum cDNA clone ddc37e03 3', mRNA sequence.
ACCESSION   BJ380110
VERSION     BJ380110.1  GI:19289494
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 471)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the culmination
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .471
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddc37e03"
                     /sex="mat A"
                     /dev_stage="Culmination stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, CF"

ORIGIN

Alignment Scores:
Pred. No.:                2.72e+03        Length:         471
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       3               Gaps:           0

US-09-858-332G-28 (1-6) x BJ380110 (1-471)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        453 CACAATCACAACCACAAC 436

RESULT 456
BQ804588
LOCUS       BQ804588                471 bp    mRNA    linear   EST 31-JUL-2002
DEFINITION  WHE3556_E02_I04ZS Wheat developing grains cDNA library Triticum
            aestivum cDNA clone WHE3556_E02_I04, mRNA sequence.
ACCESSION   BQ804588
VERSION     BQ804588.1  GI:22028751
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 471)
  AUTHORS   Altenbach,S., Anderson,O.D., Chao,S., Chin,A., Close,T.J.,
            Cronin,K., Crossman,C., Fenton,R.D., Lazo,G.R., Pham,J.,
            Rausch,C.J., Wilson,C. and Woo,J.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Developing grains cDNA library
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequences have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: SK primer.
FEATURES             Location/Qualifiers
     source          1. .471
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /cultivar="Butte 86"
                     /db_xref="taxon:4565"
                     /clone="WHE3556_E02_I04"
                     /tissue_type="whole grains"
                     /dev_stage="3-44 days post anthesis seed"
                     /lab_host="E. coli SOLR"
                     /clone_lib="Wheat developing grains cDNA library"
                     /note="Vector: Lambda ZAP II, excised phagemid; Site_1:
                     EcoRI; Plants were grown under six following different
                     environmental regimes in greenhouse, Environment 1)
                     24oC/17oC day/night, well-watered, with post-anthesis
                     fertilizer, Environment 2) 24oC/17oC day/night,
                     well-watered, without post-anthesis fertilizer,
                     Environment 3) 37oC/17oC day/night, well-watered, with
                     post-anthesis fertilizer, Environment 4) 37oC/17oC
                     day/night, well-watered, without post-anthesis fertilizer,
                     Environment 5) 37oC/17oC day/night plus drought, with
                     post-anthesis fertilizer, Environment 6) 37oC/17oC
                     day/night plus drought, without post-anthesis fertilizer,
                     developing wheat grains from the following were excised
                     and frozen in liquid nitrogen, Environment 1 at 3, 5, 7,
                     8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44 DPA Environment
                     2 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44
```

DPA Environment 3 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 34 DPA Environment 4 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 34 DPA Environment 5 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 30 DPA Environment 6 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 30 DPA and total RNA was prepared by S. Altenbach and K. Cronin at USDA-ARS, Albany, CA. A cDNA library was made using poly (A) RNA, and the cDNA clones were in vivo excised to give pBluescript SK(-) phagemids in the TJ Close lab (Chin, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (others)."

ORIGIN

Alignment Scores:

Pred. No.:	2.72e+03	Length:	471
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	5	Gaps:	0

US-09-858-332G-28 (1-6) x BQ804588 (1-471)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        330 CACAACCACAACCACAAT 347
```

RESULT 457
CF985613
LOCUS CF985613 471 bp mRNA linear EST 04-FEB-2005
DEFINITION 17031rsicee_14712.y1 Oryza sativa cv. LYP9 tillering whole plant cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA sequence.
ACCESSION CF985613
VERSION CF985613.1 GI:58587305
KEYWORDS EST.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 471)
AUTHORS Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W., Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X., Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J., Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q., Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z., Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P., Wu,S. and Liu,J.
TITLE The Genomes of Oryza sativa: A History of Duplications
JOURNAL PLoS Biol. 3 (2), e38 (2005)
PUBMED 15685292
COMMENT Contact: Yan Zhou
Bioinformatics Department
Hangzhou Genomics Institute
No.51 Zhijiang Road, Hangzhou 310008, China
Tel: 86-571-56805886
Fax: 86-571-56805884
Email: zhouyan@genomics.org.cn
Seq primer: M13 Forward
High quality sequence stop: 471
POLYA=No.

FEATURES Location/Qualifiers
source 1. .471
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/cultivar="LYP9"
/db_xref="taxon:39946"
/tissue_type="whole plant"
/dev_stage="tillering"
/clone_lib="Oryza sativa cv. LYP9 tillering whole plant cDNA library"

ORIGIN

Alignment Scores:

Pred. No.:	2.72e+03	Length:	471
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	7	Gaps:	0

US-09-858-332G-28 (1-6) x CF985613 (1-471)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        205 CACAACCACAACCACAAC 222
```

RESULT 458
BJ429835/c
LOCUS BJ429835 472 bp mRNA linear EST 13-MAR-2002
DEFINITION BJ429835 Dictyostelium discoideum cDNA library, VF Dictyostelium discoideum cDNA clone ddv5d06 3', mRNA sequence.
ACCESSION BJ429835
VERSION BJ429835.1 GI:19404557
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 472)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the vegetative stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES Location/Qualifiers
source 1. .472
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddv5d06"
/sex="mat A"
/dev_stage="Growth phase"
/clone_lib="Dictyostelium discoideum cDNA library, VF"

ORIGIN

Alignment Scores:

Pred. No.:	2.72e+03	Length:	472
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	3	Gaps:	0

US-09-858-332G-28 (1-6) x BJ429835 (1-472)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        161 CACAATCACAATCACAAT 144
```

RESULT 459
BM891644/c
LOCUS BM891644 472 bp mRNA linear EST 06-JUL-2004
DEFINITION sam42b09.y1 Gm-c1068 Glycine max cDNA clone SOYBEAN CLONE ID: Gm-c1068-7362 5' similar to TR:Q9ZRV8 Q9ZRV8 HYPOTHETICAL 8.7 KD PROTEIN. ;, mRNA sequence.
ACCESSION BM891644
VERSION BM891644.1 GI:19346764

```
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 472)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,
            Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
            Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
            Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
            Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            When it has been determined, an EST from the other end of this
            clone is listed in the 'Other ESTs on clone' field. This clone is
            available through: Biogenetic Services, 801 32nd Ave. Brookings, SD
            57006 USA (phone: 800 423 4163; email: info@biogeneticservices.com)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 428.
FEATURES             Location/Qualifiers
     source          1. .472
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /cultivar="Williams 82"
                     /db_xref="taxon:3847"
                     /clone="SOYBEAN CLONE ID: Gm-c1068-7362"
                     /tissue_type="Leaf, drought stressed, 1 month old plants,
                     greenhouse grown"
                     /lab_host="DH10B"
                     /clone_lib="Gm-c1068"
                     /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library was constructed from mRNA isolated
                     from drought stressed leaf tissue of the cultivar Williams
                     82. The month old greenhouse grown plants were deprived of
                     water for 3 days prior to harvesting the stressed leaf
                     tissue. Complementary DNA was synthesized from mRNA using
                     a primer consisting of a poly(dT) sequence with a XhoI
                     restriction site. EcoRI adapters were ligated to the
                     blunt-ended cDNA fragments followed by XhoI digestion. The
                     cDNA fragments were directionally cloned into the
                     EcoRI-XhoI restriction site of the pBluescript vector. The
                     ligated cDNA fragments were transformed into DH10B host
                     cells (GibcoBRL). This library was constructed in the
                     laboratory of Dr. Randy Shoemaker."
ORIGIN

Alignment Scores:
Pred. No.:                2.72e+03        Length:         472
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       3               Gaps:           0

US-09-858-332G-28 (1-6) x BM891644 (1-472)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        447 CATAATCATAATCATAAT 430

RESULT 460
CF192804
LOCUS       CF192804                  473 bp    mRNA    linear   EST 08-AUG-2003
DEFINITION  15c21j2.r1 Cryptococcus neoformans strain B3501 Cryptococcus
            neoformans var. neoformans cDNA clone 15c21j2 5', mRNA sequence.
ACCESSION   CF192804
VERSION     CF192804.1  GI:33514673
KEYWORDS    EST.
SOURCE      Cryptococcus neoformans var. neoformans (Filobasidiella neoformans
            var. neoformans)
  ORGANISM  Cryptococcus neoformans var. neoformans
            Eukaryota; Fungi; Basidiomycota; Hymenomycetes;
            Heterobasidiomycetes; Tremellomycetidae; Tremellales; Tremellaceae;
            Filobasidiella.
REFERENCE   1  (bases 1 to 473)
  AUTHORS   Kupfer,D.M., Drabenstot,S.D., Buchanan,K.L., Lai,H., Zhu,H.,
            Dyer,D.W., Roe,B.A. and Murphy,J.W.
  TITLE     Introns and splicing elements of five diverse fungi
  JOURNAL   Eukaryot. Cell 3 (5), 1088-1100 (2004)
  PUBMED    15470237
COMMENT     Contact: Murphy, JW
            Department of Microbiology and Immunology
            University of Oklahoma Health Sciences Center
            Oklahoma City, OK 73190, USA
            Tel: 405-271-2133 ex2133
            Email: juneann-murphy@ouhsc.edu
            This clone is available from the Fungal Genetics Stock Center,
            Phone 913-588-7044, http://www.fgsc.net/ Contact Dr. Bruce Roe
            (broe@ou.edu, www.genome.ou.edu) for sequencing questions Contact
            Dr. Juneann Murphy(juneann-murphy@ouhsc.edu) for library
            information.
            Putative full length read
            The vector to vector length is 507
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .473
                     /organism="Cryptococcus neoformans var. neoformans"
                     /mol_type="mRNA"
                     /strain="B3501"
                     /db_xref="taxon:40410"
                     /clone="15c21j2"
                     /lab_host="E. coli strain SOLR"
                     /clone_lib="Cryptococcus neoformans strain B3501"
                     /note="Vector: pBluescript sk-; Site_1: EcoRI at 5' end of
                     cDNA insert; Site_2: XhoI at 3' end of cDNA insert"
ORIGIN

Alignment Scores:
Pred. No.:                2.73e+03        Length:         473
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x CF192804 (1-473)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         27 CACAACCACAACCACAAC 44

RESULT 461
W84975
LOCUS       W84975                    473 bp    mRNA    linear   EST 12-SEP-1996
DEFINITION  mf42g01.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:407760 5' similar to gb:M38188 OVARIAN GRANULOSA CELL
            13.0 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION   W84975
VERSION     W84975.1  GI:1538607
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 473)
```

```
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1397539.
            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:251528
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 419.
FEATURES             Location/Qualifiers
     source          1. .473
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:407760"
                     /sex="unknown"
                     /tissue_type="embryo"
                     /dev_stage="13.5-14.5dpc total fetus"
                     /lab_host="DH10B"
                     /clone_lib="Soares mouse embryo NbME13.5 14.5"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTTT
                     T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                     14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                     State Univ., from 2 ]; double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector. Library went through one round of
                     normalization, and was constructed by Bento Soares and
                     M.Fatima Bonaldo.  "
ORIGIN

Alignment Scores:
Pred. No.:                2.73e+03        Length:          473
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       8               Gaps:            0

US-09-858-332G-28 (1-6) x W84975 (1-473)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        286 CACAACCATAACCACAAC 303

RESULT 462
BH357772/c
LOCUS       BH357772                473 bp    DNA     linear   GSS 03-DEC-2001
DEFINITION  CH230-95B7.TV CHORI-230 Segment 1 Rattus norvegicus genomic clone
            CH230-95B7, genomic survey sequence.
ACCESSION   BH357772
VERSION     BH357772.1  GI:17288506
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 473)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,
            Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D.,
            Riggs,F., de Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 EcoRI segment
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: CH230-95B7.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 95  row: B  column: 7
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .473
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-95B7"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 1"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
ORIGIN

Alignment Scores:
Pred. No.:                2.73e+03        Length:          473
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       9               Gaps:            0

US-09-858-332G-28 (1-6) x BH357772 (1-473)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         62 CATAATCATAATCATAAT 45

RESULT 463
BP888769
LOCUS       BP888769                474 bp    mRNA    linear   EST 16-MAR-2005
DEFINITION  BP888769 Lycopersicon esculentum maturing fruit Lycopersicon
            esculentum cDNA clone FA55AB10 5', mRNA sequence.
ACCESSION   BP888769
VERSION     BP888769.1  GI:58232240
KEYWORDS    EST.
SOURCE      Lycopersicon esculentum (Solanum lycopersicum)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 474)
  AUTHORS   Yamamoto,N., Tsugane,T., Watanabe,M., Yano,K., Maeda,F., Kuwata,C.,
            Moez,T., Nishimura,S. and Shibata,D.
  TITLE     Expressed sequence tags from the laboratory-grown miniature tomato
            (Lycopersicon esculentum) cultivar Micro-Tom and mining for single
            nucleotide polymorphisms and insertions/deletions in tomato
            cultivars
  JOURNAL   Unpublished (2005)
```

```
COMMENT     Contact: Daisuke Shibata
            Kazusa DNA Research Institute;
            2-6-7 Kazusa-Kamatari, Kisarazu, Chiba, 292-0818, Japan
            Tel: 81-438-52-3947
            Fax: 81-438-52-3948
            Email: shibata@kazusa.or.jp
            This clone was obtained at our laboratory.
            Please visit our web site
            URL:http://www.kazusa.or.jp/jsol/microtom/indexj.html (in Japanese)
            URL:http://www.kazusa.or.jp/jsol/microtom/indexj.html (in English).
FEATURES              Location/Qualifiers
     source           1. .474
                      /organism="Lycopersicon esculentum"
                      /mol_type="mRNA"
                      /cultivar="Micro-Tom"
                      /db_xref="taxon:4081"
                      /clone="FA55AB10"
                      /tissue_type="maturing fruit"
                      /clone_lib="Lycopersicon esculentum maturing fruit"
ORIGIN

Alignment Scores:
Pred. No.:              2.73e+03        Length:         474
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-858-332G-28 (1-6) x BP888769 (1-474)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        339 CACAACCACAACCACAAC 356

RESULT 464
CF962395
LOCUS       CF962395                475 bp    mRNA    linear   EST 15-JAN-2004
DEFINITION  9808siced_0882.z1 Oryza sativa cv. LYP9 trefoil whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF962395
VERSION     CF962395.1  GI:38478262
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 475)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 475
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .475
                      /organism="Oryza sativa (indica cultivar-group)"
                      /mol_type="mRNA"
                      /cultivar="LYP9"
                      /db_xref="taxon:39946"
                      /tissue_type="whole plant"
                      /dev_stage="trefoil"
                      /clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA
                      library"
ORIGIN

Alignment Scores:
Pred. No.:              2.74e+03        Length:         475
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     7               Gaps:           0

US-09-858-332G-28 (1-6) x CF962395 (1-475)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        195 CACAACCACAACCACAAC 212

RESULT 465
BZ621658/c
LOCUS       BZ621658                475 bp    DNA     linear   GSS 16-JAN-2003
DEFINITION  ig73e12.g1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
            ig73e12 5', genomic survey sequence.
ACCESSION   BZ621658
VERSION     BZ621658.1  GI:27773510
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 475)
  AUTHORS   Rabinowicz,P.D., O'Shaughnessy,A.L., Balija,V., Dedhia,N.,
            Katzenburger,F., King,L., Miller,B., Muller,S., Nascimento,L.,
            Zutavern,T., McCombie,W.R. and Martienssen,R.A.
  TITLE     Genomic shotgun sequences from Zea mays (methyl-filtered)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: W. Richard McCombie
            Lita Annenberg Hazen Genome Sequencing Center
            Cold Spring Harbor Laboratory
            PO Box 100, Cold Spring Harbor, NY 11724, USA
            Tel: 516 367 8884
            Fax: 516 367 8874
            Email: mccombie@cshl.org
            Plate: ig73  row: e  column: 12
            Seq primer: -21M13UnivRev
            Class: shotgun
            High quality sequence stop: 475.
FEATURES              Location/Qualifiers
     source           1. .475
                      /organism="Zea mays"
                      /mol_type="genomic DNA"
                      /cultivar="B73"
                      /db_xref="taxon:4577"
                      /clone="ig73e12"
                      /lab_host="DH5a"
                      /clone_lib="WGS-ZmaysF (DH5a methyl filtered)"
                      /note="Organ: immature ears; Site_1: Xba I; Site_2: Xba I;
                      The vector was digested with XbaI and one nucleotide was
                      added by fill in in the recessive 3' end. The genomic DNA
                      was nebulized, end repaired, adaptor ligated and size
                      fractionated using sephadex. The resulting fragments were
                      between 0.8 and 3 kb and were cloned into the vector (.x/y
                      reads in M13mp19, .b/g reads in pUC19). The same ligation
                      was transformed into DH5a."
ORIGIN

Alignment Scores:
```

```
Pred. No.:              2.74e+03        Length:         475
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     9               Gaps:           0

US-09-858-332G-28 (1-6) x BZ621658 (1-475)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        344 CATAACCATAACCATAAC 327

RESULT 466
AU303780/c
LOCUS       AU303780                476 bp    mRNA    linear   EST 22-JAN-2004
DEFINITION  AU303780 zinnia cultured mesophyll cell equalized cDNA Zinnia
            elegans cDNA clone Z12714, mRNA sequence.
ACCESSION   AU303780
VERSION     AU303780.1  GI:41119673
KEYWORDS    EST.
SOURCE      Zinnia elegans
  ORGANISM  Zinnia elegans
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; campanulids; Asterales; Asteraceae; Asteroideae;
            Heliantheae; Zinnia.
REFERENCE   1  (bases 1 to 476)
  AUTHORS   Demura,T., Tashiro,G., Horiguchi,G., Kishimoto,N., Kubo,M.,
            Matsuoka,N., Minami,A., Nagata-Hiwatashi,M., Nakamura,K.,
            Okamura,Y., Sassa,N., Suzuki,S., Yazaki,J., Kikuchi,S. and
            Fukuda,H.
  TITLE     Visualization by comprehensive microarray analysis of gene
            expression programs during transdifferentiation of mesophyll cells
            into xylem cells
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (24), 15794-15799 (2002)
COMMENT     Contact: Taku Demura
            Morphogenesis Research Group
            RIKEN Plant Science Center
            1-7-22 Suehirocho, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9605
            Fax: 81-45-503-9573
            Email: demura@postman.riken.go.jp
            This clone was obtained at our laboratory
            Seq primer: M13 forward.
FEATURES             Location/Qualifiers
     source          1. .476
                     /organism="Zinnia elegans"
                     /mol_type="mRNA"
                     /cultivar="Canary bird"
                     /db_xref="taxon:34245"
                     /clone="Z12714"
                     /tissue_type="mesophyll cell"
                     /clone_lib="zinnia cultured mesophyll cell equalized cDNA"
                     /note="Vector: pGEM-T easy; cultured in tracheary element
                     differentiation-inductive medium"
ORIGIN

Alignment Scores:
Pred. No.:              2.75e+03        Length:         476
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AU303780 (1-476)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        426 CATAACCATAACCATAAT 409

RESULT 467
BJ358156/c
LOCUS       BJ358156                476 bp    mRNA    linear   EST 07-MAR-2002
DEFINITION  BJ358156 Dictyostelium discoideum cDNA library, AF Dictyostelium
            discoideum cDNA clone dda65m22 3', mRNA sequence.
ACCESSION   BJ358156
VERSION     BJ358156.1  GI:19257751
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 476)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the aggregation
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .476
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dda65m22"
                     /sex="mat A"
                     /dev_stage="Aggregation stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN

Alignment Scores:
Pred. No.:              2.75e+03        Length:         476
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-858-332G-28 (1-6) x BJ358156 (1-476)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        456 CACAATCACAACCACAAC 439

RESULT 468
BJ431249/c
LOCUS       BJ431249                476 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ431249 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv13o07 3', mRNA sequence.
ACCESSION   BJ431249
VERSION     BJ431249.1  GI:19405971
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 476)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .476
```

/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddv13o07"
/sex="mat A"
/dev_stage="Growth phase"
/clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:

Pred. No.:	2.75e+03	Length:	476
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	3	Gaps:	0

US-09-858-332G-28 (1-6) x BJ431249 (1-476)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        210 CACAATCACAATCACAAT 193
```

RESULT 469
BP529011/c
LOCUS BP529011 476 bp mRNA linear EST 05-NOV-2004
DEFINITION BP529011 MAT001 Nicotiana tabacum cDNA clone BY13887, mRNA sequence.
ACCESSION BP529011
VERSION BP529011.1 GI:52832738
KEYWORDS EST.
SOURCE Nicotiana tabacum (common tobacco)
ORGANISM Nicotiana tabacum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; lamiids; Solanales; Solanaceae; Nicotiana.
REFERENCE 1 (bases 1 to 476)
AUTHORS Matsuoka,K., Demura,T., Galis,I., Horiguchi,T., Sasaki,M., Tashiro,G. and Fukuda,H.
TITLE A Comprehensive Gene Expression Analysis Toward the Understanding of Growth and Differentiation of Tobacco BY-2 Cells
JOURNAL Plant Cell Physiol. 45 (9), 1280-1289 (2004)
PUBMED 15509851
COMMENT Contact: Ken Matsuoka
Morphogenesis Research Group
RIKEN Plant Science Center
1-7-2 Suehirocho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9575
Fax: 81-45-503-9573
Email: by2@psc.riken.go.jp, URL:http://mrg.psc.riken.go.jp/strc/
The cDNA library was constructed from mRNA isolated from lag (9 h), log (72 h) and stationary (7 days) old BY-2 cells.
Seq primer: M13 forward.
FEATURES Location/Qualifiers
source 1. .476
/organism="Nicotiana tabacum"
/mol_type="mRNA"
/cultivar="Bright Yellow No.2"
/db_xref="taxon:4097"
/clone="BY13887"
/cell_line="BY-2"
/clone_lib="MAT001"
/note="Vector: pGEM-T easy; primer: M13 forward; mRNA obtained from lag, log and stationary phase cells"
ORIGIN

Alignment Scores:

Pred. No.:	2.75e+03	Length:	476
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	3	Gaps:	0

US-09-858-332G-28 (1-6) x BP529011 (1-476)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         80 CACAATCACAATCACAAT 63
```

RESULT 470
CF961968
LOCUS CF961968 476 bp mRNA linear EST 15-JAN-2004
DEFINITION 13885rsiced_16815.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA sequence.
ACCESSION CF961968
VERSION CF961968.1 GI:38477835
KEYWORDS EST.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 476)
AUTHORS Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W., Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X., Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J., Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q., Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z., Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P., Wu,S. and Liu,J.
TITLE The Genomes of Oryza sativa: A History of Duplications
JOURNAL PLoS Biol. 3 (2), e38 (2005)
PUBMED 15685292
COMMENT Contact: Yan Zhou
Bioinformatics Department
Hangzhou Genomics Institute
No.51 Zhijiang Road, Hangzhou 310008, China
Tel: 86-571-56805886
Fax: 86-571-56805884
Email: zhouyan@genomics.org.cn
Seq primer: M13 Forward
High quality sequence stop: 476
POLYA=No.
FEATURES Location/Qualifiers
source 1. .476
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/cultivar="LYP9"
/db_xref="taxon:39946"
/tissue_type="whole plant"
/dev_stage="trefoil"
/clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA library"
ORIGIN

Alignment Scores:

Pred. No.:	2.75e+03	Length:	476
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	7	Gaps:	0

US-09-858-332G-28 (1-6) x CF961968 (1-476)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        212 CACAACCACAACCACAAC 229
```

RESULT 471
CV276803/c
LOCUS CV276803 476 bp mRNA linear EST 22-SEP-2004
DEFINITION WS0141.B21_N05 PTxD-IL-A-5 Populus trichocarpa x Populus deltoides

cDNA clone WS0141_N05 3', mRNA sequence.
ACCESSION CV276803
VERSION CV276803.1 GI:52529778
KEYWORDS EST.
SOURCE Populus trichocarpa x Populus deltoides
ORGANISM Populus trichocarpa x Populus deltoides
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids I; Malpighiales; Salicaceae; Saliceae; Populus.
REFERENCE 1 (bases 1 to 476)
AUTHORS Ralph,S., Cooper,D., Kolosova,N., Oddy,C., Butterfield,Y., Kirkpatrick,R., Liu,J., Palmquist,D., Stott,J., Barber,S., Yang,G., Babakaiff,R., Brown-John,M., Chand,S., Featherstone,R., Masson,A., Mayo,M., Moran,J., Olson,T., Wong,D., Ritland,C.E., Siddiqui,A., Holt,R., Jones,S., Marra,M., Ellis,B.E., Douglas,C., Ritland,K. and Bohlmann,J.
TITLE The poplar transcriptome: Analysis of expressed sequence tags from multiple cDNA libraries
JOURNAL Unpublished (2004)
COMMENT Contact: Joerg Bohlmann
Genome BC forest genomics program
University of British Columbia
Michael Smith Laboratories, 6174 University Boulevard, Rm. 237, Vancouver, British Columbia, Canada, V6T 1Z3
Tel: 1-604-822-0282
Fax: 1-604-822-2114
Email: bohlmann@msl.ubc.ca
Plate: WS0141 row: N column: 05
High quality sequence stop: 476
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .476
/organism="Populus trichocarpa x Populus deltoides"
/mol_type="mRNA"
/cultivar="H11-11"
/db_xref="taxon:3695"
/clone="WS0141_N05"
/sex="Male"
/lab_host="E. coli DH10B T1 phage resistant cells"
/clone_lib="PTxD-IL-A-5"
/note="Vector: pBluescript II SK (+) XR; Site_1: EcoRI (5' end of cDNA); Site_2: XhoI (3' end of cDNA); Sapling trees two metres in height and grown under greenhouse conditions were exposed to continuous feeding by Malacosoma disstria Hubner (forest tent caterpillar) mid-instar larvae caged on the sapling using mesh bags. Mature leaves from within the caged region were collected 2 hours, 12 hours, 24 hours and 48 hours after the onset of treatment. mRNA was isolated from each tissue source independently and equal quantities of mRNA from each tissue were then pooled. cDNA was prepared from 5 micrograms of mRNA and directionally ligated into the pBluescript II SK (+) XR vector using the pBluescript II XR cDNA Library Construction Kit according to manufacturer's instructions with modifications (Stratagene). Plasmid DNA was then transformed by electroporation into DH10B cells (Invitrogen) for propagation."
ORIGIN

Alignment Scores:

Pred. No.:	2.75e+03	Length:	476
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	7	Gaps:	0

US-09-858-332G-28 (1-6) x CV276803 (1-476)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         73 CATAATCATAATCATAAT 56
```

RESULT 472
H09746
LOCUS H09746 476 bp mRNA linear EST 23-JUN-1995
DEFINITION ym01b07.s1 Soares infant brain 1NIB Homo sapiens cDNA clone IMAGE:46352 3', mRNA sequence.
ACCESSION H09746
VERSION H09746.1 GI:874568
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 476)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1549
High quality sequence stops: 369
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1549 Std Error: 0.00
Seq primer: Promega -21m13
High quality sequence stop: 369.
FEATURES Location/Qualifiers
source 1. .476
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:418893"
/db_xref="taxon:9606"
/clone="IMAGE:46352"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares infant brain 1NIB"
/note="Organ: whole brain; Vector: Lafmid BA; Site_1: Not I; Site_2: Hind III; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Hind III adaptors (Pharmacia), digested with Not I and directionally cloned into the Not I and Hind III sites of the Lafmid BA vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."
ORIGIN

Alignment Scores:

Pred. No.:	2.75e+03	Length:	476
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	8	Gaps:	0

US-09-858-332G-28 (1-6) x H09746 (1-476)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        200 CATAATCATAATCATAAT 217
```

RESULT 473
AQ659741/c

```
LOCUS       AQ659741                476 bp    DNA     linear   GSS 23-JUN-1999
DEFINITION  Sheared DNA-16J1.TR Sheared DNA Trypanosoma brucei genomic clone
            Sheared DNA-16J1, genomic survey sequence.
ACCESSION   AQ659741
VERSION     AQ659741.1  GI:5167509
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 476)
  AUTHORS   El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
            Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S.,
            Donelson,J., Fraser,C. and Adams,M.
  TITLE     Determination of clone end sequences from Trypanosoma brucei GUTat
            10.1 sheared DNA library
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: Sheared DNA-16J1.TF
            Contact: Najib M. El-Sayed
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: nelsayed@tigr.org
            Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
            DNA library constructed at TIGR. Clones will be available for
            distribution through ATCC. Sheared DNA end sequences search page:
            http://www.tigr.org/tdb/mdb/tbdb/.
            Seq primer: M13-Reverse
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .476
                     /organism="Trypanosoma brucei"
                     /mol_type="genomic DNA"
                     /strain="TREU927/4 GUTat 10.1"
                     /db_xref="taxon:5691"
                     /clone="Sheared DNA-16J1"
                     /clone_lib="Sheared DNA"
                     /note="Vector: pUC18; Site_1: SmaI; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from a cloned population of
                     Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
                     sheared to give a tight size distribution (approx 2 kb).
                     The v + i method used for the library construction is
                     described in detail in Smith, H.O. and Venter, J.C.
                     (Making small insert libraries for whole genome shotgun
                     sequencing projects.  In Genome Sequencing: A Practical
                     Approach, eds. M. Vaudin and B. Barell, Oxford University
                     Press, 1999)."
ORIGIN
```

Alignment Scores:

Pred. No.:	2.75e+03	Length:	476
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

US-09-858-332G-28 (1-6) x AQ659741 (1-476)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        260 CACAATCACAACCACAAC 243
```

```
RESULT 474
AA762840
LOCUS       AA762840                477 bp    mRNA    linear   EST 27-JAN-1998
DEFINITION  vs77g09.r1 Barstead mouse myotubes MPLRB5 Mus musculus cDNA clone
            IMAGE:1152352 5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0
            KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION   AA762840
VERSION     AA762840.1  GI:2812587
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 477)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:625560
            Putative full length read
            vector to vector length is 566
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 449.
FEATURES             Location/Qualifiers
     source          1. .477
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C3H"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1152352"
                     /cell_line="C2C12"
                     /lab_host="DH10B"
                     /clone_lib="Barstead mouse myotubes MPLRB5"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     [AATTCGGATCCTTG], digested with Not I and cloned into the
                     Not I  and Eco RI sites of the modified pT7T3 vector.
                     Library constructed  by Bob Barstead.  The C2C12 cell line
                     (available from ATCC,  catalog # CRL-1772) differentiates
                     rapidly, forming contractile  myotubes and producing
                     characteristic muscle proteins.  "
ORIGIN
```

Alignment Scores:

Pred. No.:	2.75e+03	Length:	477
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0

US-09-858-332G-28 (1-6) x AA762840 (1-477)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         78 CACAACCATAACCACAAC 95
```

```
RESULT 475
BP749972
LOCUS       BP749972                477 bp    mRNA    linear   EST 15-JUN-2004
DEFINITION  BP749972 partially normalized diploid tobacco cDNA library
            Nicotiana sylvestris cDNA clone R-027_D05, mRNA sequence.
ACCESSION   BP749972
VERSION     BP749972.1  GI:48759576
```

```
KEYWORDS    EST.
SOURCE      Nicotiana sylvestris (wood tobacco)
  ORGANISM  Nicotiana sylvestris
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Nicotiana.
REFERENCE   1  (bases 1 to 477)
  AUTHORS   Katoh,A., Yamaguchi,Y., Sano,H. and Hashimoto,T.
  TITLE     Analysis of expression sequence tags from Nicotiana sylvestris
  JOURNAL   Proc. Jpn. Acad. Ser. B 79, 151-154 (2003)
COMMENT     Contact: Takashi Hashimoto
            Graduate School of Biological Sciences
            Nara Institute of Science and Technology
            Takayama 8916-5, Ikoma, Nara 630-0192, Japan
            Tel: 81-743-72-5520
            Fax: 81-743-72-5529
            Email: hasimoto@bs.naist.jp.
FEATURES             Location/Qualifiers
     source          1. .477
                     /organism="Nicotiana sylvestris"
                     /mol_type="mRNA"
                     /db_xref="taxon:4096"
                     /clone="R-027_D05"
                     /tissue_type="mixture of wounded and un-wounded leaf"
                     /dev_stage="2-month-old plant"
                     /clone_lib="partially normalized diploid tobacco cDNA
                     library"
ORIGIN

Alignment Scores:
Pred. No.:              2.75e+03        Length:         477
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-858-332G-28 (1-6) x BP749972 (1-477)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        168 CACAATCACAATCACAAT 185

RESULT 476
BP878667
LOCUS       BP878667                 477 bp    mRNA    linear   EST 16-MAR-2005
DEFINITION  BP878667 Lycopersicon esculentum maturing fruit Lycopersicon
            esculentum cDNA clone FA12BH10 5', mRNA sequence.
ACCESSION   BP878667
VERSION     BP878667.1  GI:58215190
KEYWORDS    EST.
SOURCE      Lycopersicon esculentum (Solanum lycopersicum)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 477)
  AUTHORS   Yamamoto,N., Tsugane,T., Watanabe,M., Yano,K., Maeda,F., Kuwata,C.,
            Moez,T., Nishimura,S. and Shibata,D.
  TITLE     Expressed sequence tags from the laboratory-grown miniature tomato
            (Lycopersicon esculentum) cultivar Micro-Tom and mining for single
            nucleotide polymorphisms and insertions/deletions in tomato
            cultivars
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Daisuke Shibata
            Kazusa DNA Research Institute;
            2-6-7 Kazusa-Kamatari, Kisarazu, Chiba, 292-0818, Japan
            Tel: 81-438-52-3947
            Fax: 81-438-52-3948
            Email: shibata@kazusa.or.jp
            This clone was obtained at our laboratory.
            Please visit our web site
            URL:http://www.kazusa.or.jp/jsol/microtom/indexj.html (in Japanese)
            URL:http://www.kazusa.or.jp/jsol/microtom/indexj.html (in English).
FEATURES             Location/Qualifiers
     source          1. .477
                     /organism="Lycopersicon esculentum"
                     /mol_type="mRNA"
                     /cultivar="Micro-Tom"
                     /db_xref="taxon:4081"
                     /clone="FA12BH10"
                     /tissue_type="maturing fruit"
                     /clone_lib="Lycopersicon esculentum maturing fruit"
ORIGIN

Alignment Scores:
Pred. No.:              2.75e+03        Length:         477
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-858-332G-28 (1-6) x BP878667 (1-477)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        176 CACAATCACAATCACAAT 193

RESULT 477
BP879943
LOCUS       BP879943                 477 bp    mRNA    linear   EST 16-MAR-2005
DEFINITION  BP879943 Lycopersicon esculentum maturing fruit Lycopersicon
            esculentum cDNA clone FA16AE01 5', mRNA sequence.
ACCESSION   BP879943
VERSION     BP879943.1  GI:58216466
KEYWORDS    EST.
SOURCE      Lycopersicon esculentum (Solanum lycopersicum)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 477)
  AUTHORS   Yamamoto,N., Tsugane,T., Watanabe,M., Yano,K., Maeda,F., Kuwata,C.,
            Moez,T., Nishimura,S. and Shibata,D.
  TITLE     Expressed sequence tags from the laboratory-grown miniature tomato
            (Lycopersicon esculentum) cultivar Micro-Tom and mining for single
            nucleotide polymorphisms and insertions/deletions in tomato
            cultivars
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Daisuke Shibata
            Kazusa DNA Research Institute;
            2-6-7 Kazusa-Kamatari, Kisarazu, Chiba, 292-0818, Japan
            Tel: 81-438-52-3947
            Fax: 81-438-52-3948
            Email: shibata@kazusa.or.jp
            This clone was obtained at our laboratory.
            Please visit our web site
            URL:http://www.kazusa.or.jp/jsol/microtom/indexj.html (in Japanese)
            URL:http://www.kazusa.or.jp/jsol/microtom/indexj.html (in English).
FEATURES             Location/Qualifiers
     source          1. .477
                     /organism="Lycopersicon esculentum"
                     /mol_type="mRNA"
                     /cultivar="Micro-Tom"
                     /db_xref="taxon:4081"
                     /clone="FA16AE01"
                     /tissue_type="maturing fruit"
                     /clone_lib="Lycopersicon esculentum maturing fruit"
ORIGIN

Alignment Scores:
Pred. No.:              2.75e+03        Length:         477
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
```

```
Query Match:            100.00%       Indels:        0
DB:                     3             Gaps:          0

US-09-858-332G-28 (1-6) x BP879943 (1-477)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        270 CACAACCACAACCACAAC 287

RESULT 478
CD910279
LOCUS       CD910279                 478 bp    mRNA    linear   EST 14-JUL-2003
DEFINITION  G468.114O08F010821 G468 Triticum aestivum cDNA clone G468114O08,
            mRNA sequence.
ACCESSION   CD910279
VERSION     CD910279.1  GI:32684603
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 478)
  AUTHORS   Genoplante.
  TITLE     Genoplante, a major partnership french program in plant genomics
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Genoplante
            Genoplante
            93, rue Henri Rochefort 91025 EVRY CEDEX France
            Tel: 33 1 69 47 54 00
            Fax: 33 1 69 47 54 10
            This sequence has been generated in the framework of the french
            plant genomics programme 'Genoplante' (http://www.genoplante.com
            and http://genoplante-info.infobiogen.fr).
FEATURES             Location/Qualifiers
     source          1. .478
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /cultivar="recital"
                     /db_xref="taxon:4565"
                     /clone="G468114O08"
                     /tissue_type="grain (468 degrees per day after
                     pollination)"
                     /clone_lib="G468"
ORIGIN

Alignment Scores:
Pred. No.:              2.76e+03      Length:        478
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     6             Gaps:          0

US-09-858-332G-28 (1-6) x CD910279 (1-478)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        353 CACAACCACAACCACAAT 370

RESULT 479
AW224434
LOCUS       AW224434                 479 bp    mRNA    linear   EST 18-MAY-2001
DEFINITION  EST298321 tomato root deficiency, Cornell University Lycopersicon
            esculentum cDNA clone cLEW2I13, mRNA sequence.
ACCESSION   AW224434
VERSION     AW224434.1  GI:6536104
KEYWORDS    EST.
SOURCE      Lycopersicon esculentum (Solanum lycopersicum)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 479)
  AUTHORS   van der Hoeven,R.S., Garvin,D., Matern,A.L., Holt,I.E., Liang,F.,
            Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S.,
            Ronning,C.M., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and
            Tanksley,S.D.
  TITLE     Generation of ESTs from tomato root tissue
  JOURNAL   Unpublished (1999)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html
            5 prime sequence.
FEATURES             Location/Qualifiers
     source          1. .479
                     /organism="Lycopersicon esculentum"
                     /mol_type="mRNA"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cLEW2I13"
                     /tissue_type="roots"
                     /dev_stage="5-6 weeks old"
                     /clone_lib="tomato root deficiency, Cornell University"
                     /note="Vector: pBluescript SK-; Site_1: 5' EcoRI; Site_2:
                     3' XhoI; supplier: Tanksley; Tissue supplied by Dave
                     Garvin (USDA-ARS, Ithaca, NY 14850). Roots were harvested
                     from plants grown under the following
                     deficiencies/stresses: 10 mM Al, Zn, P, K, Fe, N), and
                     mRNA was isolated from indivdual treatments. Proportional
                     aliquots of mRNA of each treatment were mixed and used for
                     library construction."
ORIGIN

Alignment Scores:
Pred. No.:              2.76e+03      Length:        479
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     1             Gaps:          0

US-09-858-332G-28 (1-6) x AW224434 (1-479)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         40 CACAATCACAATCACAAT 57

RESULT 480
BQ246847
LOCUS       BQ246847                 479 bp    mRNA    linear   EST 03-MAY-2002
DEFINITION  TaE15004F05R TaE15 Triticum aestivum cDNA clone TaE15004F05R, mRNA
            sequence.
ACCESSION   BQ246847
VERSION     BQ246847.1  GI:20442723
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 479)
  AUTHORS   Cloutier,S.
  TITLE     Wheat functional genomics - Glenlea developing seeds cDNA libraries
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Dr. Sylvie Cloutier
            Cereal Research Centre, Agriculture and Agri-food Canada
            195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
            Tel: (204) 983-2340
            Fax: (204) 983-4604
            Email: scloutier@agr.gc.ca
            was cloned directionally, not all sequences generated with reverse
            primer were from the 5' end (same with forward primer and 3' end).
            Average insert size is >1.4 kb
```

```
           Plate: 004  row: F  column: 05
           Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .479
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /cultivar="Glenlea"
                     /db_xref="taxon:4565"
                     /clone="TaE15004F05R"
                     /tissue_type="developing seeds"
                     /dev_stage="15 days after anthesis"
                     /lab_host="E. coli DH10B"
                     /clone_lib="TaE15"
                     /note="Vector: pCMV-SPORT6.0 (Invitrogen Technologies);
                     Site_1: NotI; Site_2: MluI; mRNA obtained from wheat seeds
                     of cultivar Glenlea 15 days post-anthesis"
ORIGIN

Alignment Scores:
Pred. No.:              2.76e+03       Length:         479
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     3              Gaps:           0

US-09-858-332G-28 (1-6) x BQ246847 (1-479)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        297 CACAACCACAACCACAAT 314

RESULT 481
CB872086
LOCUS       CB872086                479 bp    mRNA    linear   EST 03-JUL-2003
DEFINITION  HC06I16y CH Hordeum vulgare cDNA clone HC06I16 3-PRIME, mRNA
            sequence.
ACCESSION   CB872086
VERSION     CB872086.1  GI:30074066
KEYWORDS    EST.
SOURCE      Hordeum vulgare
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 479)
  AUTHORS   Varshney,R.K., Zhang,H., Burton,R., Stein,N., Langridge,P. and
            Graner,A.
  TITLE     Barley ESTs from coleoptile tissue
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Stein Nils
            Molecular Markers Group, Department Genbank
            Institute of Plant Genetics and Crop Plant Research (IPK)
            Corrensstr. 3, 06466, Gatersleben, Germany
            Tel: 039482-5522
            Fax: 039482-5595
            Email: stein@ipk-gatersleben.de
            Insert Length: 479   Std Error: 0.00
            Plate: 6  row: I  column: 16
            Seq primer: SP6.
FEATURES             Location/Qualifiers
     source          1. .479
                     /organism="Hordeum vulgare"
                     /mol_type="mRNA"
                     /cultivar="Sloop"
                     /db_xref="GABI:550488"
                     /db_xref="taxon:4513"
                     /clone="HC06I16"
                     /tissue_type="coleoptile"
                     /dev_stage="coleoptile, 3 days old"
                     /lab_host="DH10B"
                     /clone_lib="CH"
                     /note="Vector: pSPORT; Site_1: SalI (5-end of cDNA);
                     Site_2: NotI (3-end of cDNA); Due to the cloning system
                     used blue/white selection for recombinants is not 100%
                     reliable. Average insert size is 1.3 kb."
ORIGIN

Alignment Scores:
Pred. No.:              2.76e+03       Length:         479
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     6              Gaps:           0

US-09-858-332G-28 (1-6) x CB872086 (1-479)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        299 CATAACCATAATCATAAC 316

RESULT 482
CF193670
LOCUS       CF193670                479 bp    mRNA    linear   EST 08-AUG-2003
DEFINITION  l6107j2.r1 Cryptococcus neoformans strain B3501 Cryptococcus
            neoformans var. neoformans cDNA clone l6107j2 5', mRNA sequence.
ACCESSION   CF193670
VERSION     CF193670.1  GI:33515553
KEYWORDS    EST.
SOURCE      Cryptococcus neoformans var. neoformans (Filobasidiella neoformans
            var. neoformans)
  ORGANISM  Cryptococcus neoformans var. neoformans
            Eukaryota; Fungi; Basidiomycota; Hymenomycetes;
            Heterobasidiomycetes; Tremellomycetidae; Tremellales; Tremellaceae;
            Filobasidiella.
REFERENCE   1  (bases 1 to 479)
  AUTHORS   Kupfer,D.M., Drabenstot,S.D., Buchanan,K.L., Lai,H., Zhu,H.,
            Dyer,D.W., Roe,B.A. and Murphy,J.W.
  TITLE     Introns and splicing elements of five diverse fungi
  JOURNAL   Eukaryot. Cell 3 (5), 1088-1100 (2004)
   PUBMED   15470237
COMMENT     Other_ESTs: l6107j2.f1
            Contact: Murphy, JW
            Department of Microbiology and Immunology
            University of Oklahoma Health Sciences Center
            Oklahoma City, OK 73190, USA
            Tel: 405-271-2133 ex2133
            Email: juneann-murphy@ouhsc.edu
            This clone is available from the Fungal Genetics Stock Center,
            Phone 913-588-7044, http://www.fgsc.net/ Contact Dr. Bruce Roe
            (broe@ou.edu, www.genome.ou.edu) for sequencing questions Contact
            Dr. Juneann Murphy(juneann-murphy@ouhsc.edu) for library
            information.
            Putative full length read
            The vector to vector length is 503
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .479
                     /organism="Cryptococcus neoformans var. neoformans"
                     /mol_type="mRNA"
                     /strain="B3501"
                     /db_xref="taxon:40410"
                     /clone="l6107j2"
                     /lab_host="E. coli strain SOLR"
                     /clone_lib="Cryptococcus neoformans strain B3501"
                     /note="Vector: pBluescript sk-; Site_1: EcoRI at 5' end of
                     cDNA insert; Site_2: XhoI at 3' end of cDNA insert"
ORIGIN

Alignment Scores:
Pred. No.:              2.76e+03       Length:         479
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
```

```
DB:                    6              Gaps:          0

US-09-858-332G-28 (1-6) x CF193670 (1-479)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         35 CACAACCACAACCACAAC 52

RESULT 483
PT013G05U/c
LOCUS       PT013G05U               479 bp    DNA     linear   GSS 29-MAY-2003
DEFINITION  Paramecium tetraurelia sequence R13D03u of the end of plasmid
            PT013G05, genomic survey sequence.
ACCESSION   AL446113
VERSION     AL446113.1  GI:11123000
KEYWORDS    GSS.
SOURCE      Paramecium tetraurelia
  ORGANISM  Paramecium tetraurelia
            Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
            Paramecium.
REFERENCE   1
  AUTHORS   Keller,A.M. and Cohen,J.
  TITLE     An indexed genomic library for Paramecium complementation cloning
  JOURNAL   J. Eukaryot. Microbiol. 47 (1), 1-6 (2000)
   PUBMED   10651287
REFERENCE   2  (bases 1 to 479)
  AUTHORS   Dessen,P., Zagulski,M., Gromadka,R., Plattner,H., Kissmehl,R.,
            Meyer,E., Betermier,M., Schultz,J.E., Linder,J.U., Pearlman,R.E.,
            Kung,C., Forney,J., Satir,B.H., Van Houten,J.L., Keller,A.M.,
            Froissard,M., Sperling,L. and Cohen,J.
  TITLE     Paramecium genome survey: a pilot project
  JOURNAL   Trends Genet. 17 (6), 306-308 (2001)
   PUBMED   11377780
REFERENCE   3  (bases 1 to 479)
  AUTHORS   Wong,L. and Pearlman,R.E.
  TITLE     Random sequencing of the Paramecium macronuclear genome
  JOURNAL   Unpublished
  REMARK    Core Molecular Biology Facility, York University, 4700 Keele St.,
            Toronto, Canada M3J 1P3
REFERENCE   4  (bases 1 to 479)
  AUTHORS   Cohen,J. and Sperling,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-NOV-2000) Paramecium Genome Survey Project, Centre de
            Genetique Moleculaire, Centre National de la Recherche
            Scientifique, 91198 Gif-sur-Yvette, France. E-mail:
            sperling@cgm.cnrs-gif.fr
COMMENT     The present survey of the Paramecium tetraurelia macronuclear
            genome consists of end sequences of a library of random 4-12 kb
            fragments obtained by Sau3A partial digestion of macronuclear DNA
            cloned in the BamHI site of pBSIIKS-. See [4].
            Genes are predicted from matches to other sequences. For more
            information about this sequence or the Paramecium Project, see
            http://paramecium.cgm.cnrs-gif.fr.
FEATURES             Location/Qualifiers
     source          1. .479
                     /organism="Paramecium tetraurelia"
                     /macronuclear
                     /mol_type="genomic DNA"
                     /strain="stock d4-2"
                     /db_xref="taxon:5888"
ORIGIN

Alignment Scores:
Pred. No.:              2.76e+03        Length:         479
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     11              Gaps:           0

US-09-858-332G-28 (1-6) x PT013G05U (1-479)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        345 CATAATCATAATCATAAT 328

RESULT 484
AA667520
LOCUS       AA667520                480 bp    mRNA    linear   EST 19-NOV-1997
DEFINITION  vv18a05.r1 Stratagene mouse heart (#937316) Mus musculus cDNA clone
            IMAGE:1222736 5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0
            KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION   AA667520
VERSION     AA667520.1  GI:2626221
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 480)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:648328
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 431.
FEATURES             Location/Qualifiers
     source          1. .480
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="NIH Swiss"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1222736"
                     /sex="pooled"
                     /tissue_type="heart"
                     /dev_stage="13 day embryos"
                     /lab_host="SOLR (kanamycin resistant)"
                     /clone_lib="Stratagene mouse heart (#937316)"
                     /note="Organ: heart; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. 93 pooled NIH/Swiss 13 day embryo hearts.
                     Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5'
                     adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
ORIGIN

Alignment Scores:
Pred. No.:              2.77e+03        Length:         480
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AA667520 (1-480)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        224 CACAACCATAACCACAAC 241

RESULT 485
AV423668
```

```
LOCUS       AV423668                480 bp    mRNA    linear   EST 10-MAR-2004
DEFINITION  AV423668 Lotus japonicus young plants (two-week old) Lotus
            corniculatus var. japonicus cDNA clone MWM030e08_r 5', mRNA
            sequence.
ACCESSION   AV423668
VERSION     AV423668.1  GI:7779812
KEYWORDS    EST.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1  (bases 1 to 480)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Generation of 7137 non-redundant expressed sequence tags from a
            legume, Lotus japonicus
  JOURNAL   DNA Res. 7 (2), 127-130 (2000)
   PUBMED   10819328
COMMENT     Contact: Yasukazu Nakamura
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: ynakamu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .480
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="mRNA"
                     /isolate="Miyakojima MG-20"
                     /db_xref="taxon:34305"
                     /clone="MWM030e08_r"
                     /tissue_type="whole plant"
                     /dev_stage="two-week old plant"
                     /clone_lib="Lotus japonicus young plants (two-week old)"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI; synonym: Lotus japonicus"
ORIGIN

Alignment Scores:
Pred. No.:                2.77e+03        Length:         480
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       1               Gaps:           0

US-09-858-332G-28 (1-6) x AV423668 (1-480)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db          7 CACAACCACAACCACAAC 24

RESULT 486
BF293843
LOCUS       BF293843                480 bp    mRNA    linear   EST 17-NOV-2000
DEFINITION  WHE2153_F06_L11ZS Triticum turgidum L. var. durum (durum wheat)
            whole plant cDNA library Triticum turgidum cDNA clone
            WHE2153_F06_L11, mRNA sequence.
ACCESSION   BF293843
VERSION     BF293843.1  GI:11224907
KEYWORDS    EST.
SOURCE      Triticum turgidum
  ORGANISM  Triticum turgidum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 480)
  AUTHORS   Akhunov,E., Anderson,O.D., Chao,S., Chin,A., Choi,D.W., Close,T.J.,
            Fenton,R.D., Han,P.S., Hsia,C.C., Kang,Y., Kianian,P., Lazo,G.R.,
            Miller,R., Otto,C., Rausch,C.J., Seaton,C.L., Simons,K., Tong,J.C.
            and Zhang,D.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Whole plant cDNA library from Triticum turgidum L. var.
            durum
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene SK primer.
FEATURES             Location/Qualifiers
     source          1. .480
                     /organism="Triticum turgidum"
                     /mol_type="mRNA"
                     /cultivar="Langdon-16"
                     /db_xref="taxon:4571"
                     /clone="WHE2153_F06_L11"
                     /tissue_type="All tissues"
                     /dev_stage="Different growth stages"
                     /lab_host="E. coli SOLR"
                     /clone_lib="Triticum turgidum L. var. durum (durum wheat)
                     whole plant cDNA library"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; Plants were grown in a growth
                     chamber at North Dakota State University (Kianian, Otto,
                     Simons). Tissues collected from seven-day etiolated
                     seedling leaf, stem, root and seed; leaf from plant at
                     fourth leaf stage; spike from pre-anthesis through 20
                     days after anthesis; flag leaf; leaf and stem tissue from
                     tillers. and root. Total RNA and poly(A) RNA were
                     prepared from each tissue and then pooled, a cDNA library
                     was made, and the cDNA clones were in vivo excised to give
                     pBluescript phagemids in the TJ Close lab (Akhunov, Chin,
                     Choi, Close, Fenton, Kianian, Otto, Simons, Zhang) at the
                     University of California, Riverside. Plasmid DNA
                     preparations and DNA sequencing were performed in the OD
                     Anderson lab (all other authors)."
ORIGIN

Alignment Scores:
Pred. No.:                2.77e+03        Length:         480
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       2               Gaps:           0

US-09-858-332G-28 (1-6) x BF293843 (1-480)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        312 CACAACCACAACCACAAT 329

RESULT 487
BP080661
LOCUS       BP080661                480 bp    mRNA    linear   EST 27-AUG-2004
DEFINITION  BP080661 Lotus japonicus roots Lotus corniculatus var. japonicus
            cDNA clone MR061c02_f 3', mRNA sequence.
ACCESSION   BP080661
VERSION     BP080661.1  GI:45637322
KEYWORDS    EST.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1  (bases 1 to 480)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Characteristics of the Lotus japonicus gene repertoire deduced from
            large-scale expressed sequence tag (EST) analysis
```

```
  JOURNAL   Plant Mol. Biol. 54 (3), 405-414 (2004)
   PUBMED   15284495
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .480
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="mRNA"
                     /isolate="Miyakojima MG-20"
                     /db_xref="taxon:34305"
                     /clone="MR061c02_f"
                     /tissue_type="roots"
                     /clone_lib="Lotus japonicus roots"
ORIGIN

Alignment Scores:
Pred. No.:              2.77e+03        Length:         480
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-858-332G-28 (1-6) x BP080661 (1-480)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        149 CATAATCATAATCATAAT 166

RESULT 488
CO271880
LOCUS       CO271880                480 bp    mRNA    linear   EST 24-JUN-2004
DEFINITION  EK105028.5prime Exelixis FlyTag CK01 pCDNA-SK+ Drosophila
            melanogaster cDNA clone EK105028 5, mRNA sequence.
ACCESSION   CO271880
VERSION     CO271880.1  GI:49192899
KEYWORDS    EST.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 480)
  AUTHORS   Kopczynski,C., Platt,D., Campbell,J., Muzong,C., Laufer,A.,
            Peterson,E. and Swimmer,C.
  TITLE     Exelixis FlyTag EST Project CK01 Library
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: EK.1050  row: C  column: 4
            High quality sequence stop: 475.
FEATURES             Location/Qualifiers
     source          1. .480
                     /organism="Drosophila melanogaster"
                     /mol_type="mRNA"
                     /db_xref="taxon:7227"
                     /clone="EK105028"
                     /clone_lib="Exelixis FlyTag CK01 pCDNA-SK+"
                     /note="Organ: mixed stage embryos, imaginal disks, and
                     adult heads; Vector: pCDNA-SK+; Site_1: NotI; Site_2:
                     XhoI; Random primed, normalized library from mixed stage
                     embryos, imaginal disks, and adult heads."
ORIGIN

Alignment Scores:
Pred. No.:              2.77e+03        Length:         480
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     7               Gaps:           0

US-09-858-332G-28 (1-6) x CO271880 (1-480)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         83 CATAATCATAATCATAAT 100

RESULT 489
CR294373/c
LOCUS       CR294373                480 bp    DNA     linear   GSS 19-NOV-2004
DEFINITION  mte1-10O13FM1 BAC end, cultivar Jemalong A17 of Medicago
            truncatula, genomic survey sequence.
ACCESSION   CR294373
VERSION     CR294373.1  GI:44684797
KEYWORDS    GSS.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 480)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
FEATURES             Location/Qualifiers
     source          1. .480
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /cultivar="Jemalong A17"
                     /db_xref="taxon:3880"
                     /clone_lib="MTE1"
                     /note="Vector: pIndigoBAC ; Site_1: EcoRI ; Site_2: EcoRI
                     ; Debelle F. and Chalhoub B.
                     Genoscope sequence ID : mte1-10O13FM1"
ORIGIN

Alignment Scores:
Pred. No.:              2.77e+03        Length:         480
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     11              Gaps:           0

US-09-858-332G-28 (1-6) x CR294373 (1-480)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        274 CACAATCATAATCATAAT 257

RESULT 490
AI992710
LOCUS       AI992710                481 bp    mRNA    linear   EST 08-SEP-1999
DEFINITION  701493365 A. thaliana, Ohio State clone set Arabidopsis thaliana
            cDNA clone 701493365, mRNA sequence.
ACCESSION   AI992710
VERSION     AI992710.1  GI:5839615
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 481)
  AUTHORS   Chen,J., Momiyama,M., Chan,E., Mooney,M., Carroon,B., Gilliland,D.,
```

```
            Wang,X., Hillman,J., Guegler,K., Kim,C., Doyle,M., Brzoska,P.,
            Gorgone,G., Burns,D., Griffin,J., Mouanoutoua,M., Nguyen,D.,
            Tan,R., Rose,M., Warren,B., Ton,B., Kastury,K., Borillo,C.,
            Carpio,T., Policky,J., Suzuki,G., Argentine,C., Shah,S.,
            Nobriga,A., Murry,L., Turner,C., Krikorian,S., Elder,L. and
            Hanson,D.
  TITLE     Arabidopsis thaliana Gene Expression MicroArray
  JOURNAL   Unpublished (1999)
COMMENT     Contact: David Smoller, Ph.D.
            Genome Systems, Inc., a wholly owned subsidiary of Incyte
            Pharmaceuticals, Inc.
            4633 World Parkway Circle, St. Louis, MO 63134, USA
            Tel: 877-577-2733
            Fax: 314-427-3324
            Email: service@genomesystems.com.
FEATURES             Location/Qualifiers
     source          1. .481
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="701493365"
                     /clone_lib="A. thaliana, Ohio State clone set"
                     /note="cDNA library was made from selected clones from the
                     Arabidopsis thaliana Ohio  State clone set."
ORIGIN

Alignment Scores:
Pred. No.:              2.78e+03        Length:         481
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AI992710 (1-481)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        296 CACAATCATAATCACAAT 313

RESULT 491
BJ435882/c
LOCUS       BJ435882                481 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ435882 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv28e23 3', mRNA sequence.
ACCESSION   BJ435882
VERSION     BJ435882.1  GI:19410604
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 481)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .481
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv28e23"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:              2.78e+03        Length:         481
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-858-332G-28 (1-6) x BJ435882 (1-481)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        346 CACAATCACAATCACAAT 329

RESULT 492
BP888020
LOCUS       BP888020                481 bp    mRNA    linear   EST 16-MAR-2005
DEFINITION  BP888020 Lycopersicon esculentum maturing fruit Lycopersicon
            esculentum cDNA clone FA52BG07 5', mRNA sequence.
ACCESSION   BP888020
VERSION     BP888020.1  GI:58231491
KEYWORDS    EST.
SOURCE      Lycopersicon esculentum (Solanum lycopersicum)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 481)
  AUTHORS   Yamamoto,N., Tsugane,T., Watanabe,M., Yano,K., Maeda,F., Kuwata,C.,
            Moez,T., Nishimura,S. and Shibata,D.
  TITLE     Expressed sequence tags from the laboratory-grown miniature tomato
            (Lycopersicon esculentum) cultivar Micro-Tom and mining for single
            nucleotide polymorphisms and insertions/deletions in tomato
            cultivars
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Daisuke Shibata
            Kazusa DNA Research Institute;
            2-6-7 Kazusa-Kamatari, Kisarazu, Chiba, 292-0818, Japan
            Tel: 81-438-52-3947
            Fax: 81-438-52-3948
            Email: shibata@kazusa.or.jp
            This clone was obtained at our laboratory.
            Please visit our web site
            URL:http://www.kazusa.or.jp/jsol/microtom/indexj.html (in Japanese)
            URL:http://www.kazusa.or.jp/jsol/microtom/indexj.html (in English).
FEATURES             Location/Qualifiers
     source          1. .481
                     /organism="Lycopersicon esculentum"
                     /mol_type="mRNA"
                     /cultivar="Micro-Tom"
                     /db_xref="taxon:4081"
                     /clone="FA52BG07"
                     /tissue_type="maturing fruit"
                     /clone_lib="Lycopersicon esculentum maturing fruit"
ORIGIN

Alignment Scores:
Pred. No.:              2.78e+03        Length:         481
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-858-332G-28 (1-6) x BP888020 (1-481)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        380 CACAATCACAATCACAAT 397

RESULT 493
```

```
BP888943
LOCUS       BP888943                481 bp    mRNA    linear   EST 16-MAR-2005
DEFINITION  BP888943 Lycopersicon esculentum maturing fruit Lycopersicon
            esculentum cDNA clone FA55CD01 5', mRNA sequence.
ACCESSION   BP888943
VERSION     BP888943.1  GI:58232414
KEYWORDS    EST.
SOURCE      Lycopersicon esculentum (Solanum lycopersicum)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 481)
  AUTHORS   Yamamoto,N., Tsugane,T., Watanabe,M., Yano,K., Maeda,F., Kuwata,C.,
            Moez,T., Nishimura,S. and Shibata,D.
  TITLE     Expressed sequence tags from the laboratory-grown miniature tomato
            (Lycopersicon esculentum) cultivar Micro-Tom and mining for single
            nucleotide polymorphisms and insertions/deletions in tomato
            cultivars
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Daisuke Shibata
            Kazusa DNA Research Institute;
            2-6-7 Kazusa-Kamatari, Kisarazu, Chiba, 292-0818, Japan
            Tel: 81-438-52-3947
            Fax: 81-438-52-3948
            Email: shibata@kazusa.or.jp
            This clone was obtained at our laboratory.
            Please visit our web site
            URL:http://www.kazusa.or.jp/jsol/microtom/indexj.html (in Japanese)
            URL:http://www.kazusa.or.jp/jsol/microtom/indexj.html (in English).
FEATURES             Location/Qualifiers
     source          1. .481
                     /organism="Lycopersicon esculentum"
                     /mol_type="mRNA"
                     /cultivar="Micro-Tom"
                     /db_xref="taxon:4081"
                     /clone="FA55CD01"
                     /tissue_type="maturing fruit"
                     /clone_lib="Lycopersicon esculentum maturing fruit"
ORIGIN

Alignment Scores:
Pred. No.:                2.78e+03        Length:          481
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       3               Gaps:            0

US-09-858-332G-28 (1-6) x BP888943 (1-481)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        343 CACAACCACAACCACAAC 360

RESULT 494
CF988072
LOCUS       CF988072                481 bp    mRNA    linear   EST 04-FEB-2005
DEFINITION  16952rsicee_14621.y1 Oryza sativa cv. LYP9 tillering whole plant
            cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF988072
VERSION     CF988072.1  GI:58589764
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 481)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
  PUBMED    15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 481
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .481
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="tillering"
                     /clone_lib="Oryza sativa cv. LYP9 tillering whole plant
                     cDNA library"
ORIGIN

Alignment Scores:
Pred. No.:                2.78e+03        Length:          481
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       7               Gaps:            0

US-09-858-332G-28 (1-6) x CF988072 (1-481)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        218 CACAACCACAACCACAAC 235

RESULT 495
AJ776664
LOCUS       AJ776664                482 bp    mRNA    linear   EST 06-AUG-2004
DEFINITION  AJ776664 Populus euphratica root 3-6 months Populus euphratica cDNA
            clone P0000400016F12F1, mRNA sequence.
ACCESSION   AJ776664
VERSION     AJ776664.1  GI:50068339
KEYWORDS    EST.
SOURCE      Populus euphratica
  ORGANISM  Populus euphratica
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Malpighiales; Salicaceae; Saliceae; Populus.
REFERENCE   1  (bases 1 to 482)
  AUTHORS   Brosche,M., Alatalo,E.R., Vinocur,B., Altman,A., Teichmann,T.,
            Ottow,E.A., Polle,A., Djilianov,D., Afif,D., Triboulot,M.B.,
            Dreyer,E., Paulin,L. and Kangasjaervi,J.
  TITLE     Gene expression in desert leaves
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Prof. Jaakko Kangasjarvi
            Department of Biosciences, Plant Physiology
            University of Helsinki
            FIN-00014 Helsinki, Finland.
FEATURES             Location/Qualifiers
     source          1. .482
                     /organism="Populus euphratica"
                     /mol_type="mRNA"
                     /db_xref="taxon:75702"
                     /clone="P0000400016F12F1"
                     /tissue_type="root"
```

```
                     /dev_stage="3-6 months"
                     /clone_lib="Populus euphratica root 3-6 months"
                     /note="country: China:Xinjiang"
ORIGIN

Alignment Scores:
Pred. No.:                2.78e+03       Length:         482
Score:                    42.00          Matches:        6
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       1              Gaps:           0

US-09-858-332G-28 (1-6) x AJ776664 (1-482)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         63 CACAATCACAATCACAAT 80

RESULT 496
AV422803
LOCUS       AV422803                482 bp    mRNA    linear   EST 10-MAR-2004
DEFINITION  AV422803 Lotus japonicus young plants (two-week old) Lotus
            corniculatus var. japonicus cDNA clone MWM018c10_r 5', mRNA
            sequence.
ACCESSION   AV422803
VERSION     AV422803.1  GI:7778073
KEYWORDS    EST.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1  (bases 1 to 482)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Generation of 7137 non-redundant expressed sequence tags from a
            legume, Lotus japonicus
  JOURNAL   DNA Res. 7 (2), 127-130 (2000)
   PUBMED   10819328
COMMENT     Contact: Yasukazu Nakamura
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: ynakamu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .482
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="mRNA"
                     /isolate="Miyakojima MG-20"
                     /db_xref="taxon:34305"
                     /clone="MWM018c10_r"
                     /tissue_type="whole plant"
                     /dev_stage="two-week old plant"
                     /clone_lib="Lotus japonicus young plants (two-week old)"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI; synonym: Lotus japonicus"
ORIGIN

Alignment Scores:
Pred. No.:                2.78e+03       Length:         482
Score:                    42.00          Matches:        6
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       1              Gaps:           0

US-09-858-332G-28 (1-6) x AV422803 (1-482)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         57 CACAACCACAACCACAAC 74

RESULT 497
CB881699
LOCUS       CB881699                482 bp    mRNA    linear   EST 23-APR-2003
DEFINITION  HM10J21w HM Hordeum vulgare subsp. vulgare cDNA clone HM10J21
            3-PRIME, mRNA sequence.
ACCESSION   CB881699
VERSION     CB881699.1  GI:30083692
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 482)
  AUTHORS   Zhang,H., Weschke,W., Michalek,W., Stein,N. and Graner,A.
  TITLE     EST sequencing and analysis in barley (2002)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Stein Nils
            Molecular Markers Group, Department Genbank
            Institute of Plant Genetics and Crop Plant Research (IPK)
            Correnssstr. 3, 06466, Gatersleben, Germany
            Tel: 039482-5522
            Fax: 039482-5595
            Email: stein@ipk-gatersleben.de
            Insert Length: 482   Std Error: 0.00
            Plate: 10  row: J  column: 21
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .482
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Barke"
                     /sub_species="vulgare"
                     /db_xref="GABI:560245"
                     /db_xref="taxon:112509"
                     /clone="HM10J21"
                     /tissue_type="male inflorescences"
                     /dev_stage="male inflorescences (approx. 2 mm in size),
                     green anther stage"
                     /lab_host="XL10-Gold"
                     /clone_lib="HM"
                     /note="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of
                     cDNA); Site_2: XhoI (3'-end of cDNA); Due to a cloning
                     artefact caused by the kit, in most cases the EcoRI site
                     is NOT present, as well as the EcoRIadapter used for
                     cloning. To excise the insert, restriction sites upstream
                     EcoRI should be used (e.g. BamHI, SalI,PstI). NOTE: Also
                     due to the cloning system used Blue/white selection for
                     recombinats is not 100% reliable."
ORIGIN

Alignment Scores:
Pred. No.:                2.78e+03       Length:         482
Score:                    42.00          Matches:        6
Percent Similarity:       100.00%        Conservative:   0
Best Local Similarity:    100.00%        Mismatches:     0
Query Match:              100.00%        Indels:         0
DB:                       6              Gaps:           0

US-09-858-332G-28 (1-6) x CB881699 (1-482)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        349 CATAACCATAATCATAAC 366

RESULT 498
BH833860/c
LOCUS       BH833860                482 bp    DNA     linear   GSS 20-MAY-2002
DEFINITION  BACPP5-D02.y Pristionchus pacificus BAC ends Pristionchus pacificus
            genomic, genomic survey sequence.
ACCESSION   BH833860
VERSION     BH833860.1  GI:21034122
KEYWORDS    GSS.
```

```
SOURCE      Pristionchus pacificus
  ORGANISM  Pristionchus pacificus
            Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
            Neodiplogasteridae; Pristionchus.
REFERENCE   1  (bases 1 to 482)
  AUTHORS   Srinivasan,J., Sinz,W., Lanz,C., Brand,A., Nandakumar,R.,
            Raddatz,G., Witte,H., Keller,H., Kipping,I., Pires da Silva,A.,
            Jesse,T., Millare,J., de Both,M., Schuster,S.C. and Sommer,R.J.
  TITLE     A BAC-based genetic linkage map of the nematode Pristionchus
            pacificus
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Sommer RJ
            Evolutionary Biology
            Max-Planck-Institute for Developmental Biology
            Spemannstr. 37-39, Tuebingen D-72076, Germany
            Tel: 00497071601371
            Fax: 00497071601498
            Email: ralf.sommer@tuebingen.mpg.de
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .482
                     /organism="Pristionchus pacificus"
                     /mol_type="genomic DNA"
                     /strain="var. California"
                     /db_xref="taxon:54126"
                     /clone_lib="Pristionchus pacificus BAC ends"
ORIGIN

Alignment Scores:
Pred. No.:                2.78e+03        Length:         482
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       9               Gaps:           0

US-09-858-332G-28 (1-6) x BH833860 (1-482)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        320 CACAATCATAATCATAAT 303

RESULT 499
AV537689/c
LOCUS       AV537689                483 bp    mRNA    linear   EST 20-FEB-2004
DEFINITION  AV537689 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
            cDNA clone RZ101c05F 3', mRNA sequence.
ACCESSION   AV537689
VERSION     AV537689.1  GI:8697972
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 483)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     A large scale analysis of cDNA in Arabidopsis thaliana: Generation
            of 12,028 non-redundant expressed sequence tags from normalized and
            size-selected cDNA libraries
  JOURNAL   DNA Res. 7 (3), 175-180 (2000)
   PUBMED   10907847
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .483
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /ecotype="Columbia"
                     /db_xref="taxon:3702"
                     /clone="RZ101c05F"
                     /tissue_type="roots"
                     /clone_lib="Arabidopsis thaliana roots Columbia"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
ORIGIN

Alignment Scores:
Pred. No.:                2.79e+03        Length:         483
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       1               Gaps:           0

US-09-858-332G-28 (1-6) x AV537689 (1-483)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        284 CACAACCACAACCACAAC 267

RESULT 500
CB866638/c
LOCUS       CB866638                483 bp    mRNA    linear   EST 03-JUL-2003
DEFINITION  HC06I16w CH Hordeum vulgare cDNA clone HC06I16 5-PRIME, mRNA
            sequence.
ACCESSION   CB866638
VERSION     CB866638.1  GI:30068614
KEYWORDS    EST.
SOURCE      Hordeum vulgare
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 483)
  AUTHORS   Varshney,R.K., Zhang,H., Burton,R., Stein,N., Langridge,P. and
            Graner,A.
  TITLE     Barley ESTs from coleoptile tissue
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Stein Nils
            Molecular Markers Group, Department Genbank
            Institute of Plant Genetics and Crop Plant Research (IPK)
            Correnssstr. 3, 06466, Gatersleben, Germany
            Tel: 039482-5522
            Fax: 039482-5595
            Email: stein@ipk-gatersleben.de
            Insert Length: 483   Std Error: 0.00
            Plate: 6  row: I  column: 16
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .483
                     /organism="Hordeum vulgare"
                     /mol_type="mRNA"
                     /cultivar="Sloop"
                     /db_xref="GABI:532421"
                     /db_xref="taxon:4513"
                     /clone="HC06I16"
                     /tissue_type="coleoptile"
                     /dev_stage="coleoptile, 3 days old"
                     /lab_host="DH10B"
                     /clone_lib="CH"
                     /note="Vector: pSPORT; Site_1: SalI (5-end of cDNA);
                     Site_2: NotI (3-end of cDNA); Due to the cloning system
                     used blue/white selection for recombinants is not 100%
                     reliable. Average insert size is 1.3 kb."
ORIGIN

Alignment Scores:
Pred. No.:                2.79e+03        Length:         483
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
```

```
DB:                    6              Gaps:          0

US-09-858-332G-28 (1-6) x CB866638 (1-483)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        190 CATAACCATAATCATAAC 173

RESULT 501
BF293358
LOCUS       BF293358                 484 bp    mRNA    linear   EST 17-NOV-2000
DEFINITION  WHE2157_A02_B03ZS Triticum turgidum L. var. durum (durum wheat)
            whole plant cDNA library Triticum turgidum cDNA clone
            WHE2157_A02_B03, mRNA sequence.
ACCESSION   BF293358
VERSION     BF293358.1  GI:11224518
KEYWORDS    EST.
SOURCE      Triticum turgidum
  ORGANISM  Triticum turgidum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 484)
  AUTHORS   Akhunov,E., Anderson,O.D., Chao,S., Chin,A., Choi,D.W., Close,T.J.,
            Fenton,R.D., Han,P.S., Hsia,C.C., Kang,Y., Kianian,P., Lazo,G.R.,
            Miller,R., Otto,C., Rausch,C.J., Seaton,C.L., Simons,K., Tong,J.C.
            and Zhang,D.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Whole plant cDNA library from Triticum turgidum L. var.
            durum
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene SK primer.
FEATURES             Location/Qualifiers
     source          1. .484
                     /organism="Triticum turgidum"
                     /mol_type="mRNA"
                     /cultivar="Langdon-16"
                     /db_xref="taxon:4571"
                     /clone="WHE2157_A02_B03"
                     /tissue_type="All tissues"
                     /dev_stage="Different growth stages"
                     /lab_host="E. coli SOLR"
                     /clone_lib="Triticum turgidum L. var. durum (durum wheat)
                     whole plant cDNA library"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; Plants were grown in a growth
                     chamber at North Dakota State University  (Kianian, Otto,
                     Simons).  Tissues collected from seven-day etiolated
                     seedling leaf, stem, root and seed; leaf from plant at
                     fourth leaf  stage; spike from pre-anthesis through 20
                     days after anthesis; flag  leaf; leaf and stem tissue from
                     tillers. and root.  Total RNA and  poly(A) RNA were
                     prepared from each tissue and then pooled, a cDNA  library
                     was made, and the cDNA clones were in vivo excised to give
                     pBluescript phagemids in the TJ Close lab (Akhunov, Chin,
                     Choi, Close,  Fenton, Kianian, Otto, Simons, Zhang) at the
                     University of California,  Riverside.  Plasmid DNA
                     preparations and DNA sequencing were performed  in the OD
                     Anderson lab (all other authors)."
ORIGIN

Alignment Scores:
Pred. No.:             2.8e+03        Length:        484
Score:                 42.00          Matches:       6
Percent Similarity:    100.00%        Conservative:  0
Best Local Similarity: 100.00%        Mismatches:    0
Query Match:           100.00%        Indels:        0
DB:                    2              Gaps:          0

US-09-858-332G-28 (1-6) x BF293358 (1-484)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        329 CACAACCACAACCACAAT 346

RESULT 502
BJ369193
LOCUS       BJ369193                 484 bp    mRNA    linear   EST 08-MAR-2002
DEFINITION  BJ369193 Dictyostelium discoideum cDNA library, CF Dictyostelium
            discoideum cDNA clone ddc49m14 5', mRNA sequence.
ACCESSION   BJ369193
VERSION     BJ369193.1  GI:19278576
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 484)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the culmination
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .484
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddc49m14"
                     /sex="mat A"
                     /dev_stage="Culmination stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, CF"
ORIGIN

Alignment Scores:
Pred. No.:             2.8e+03        Length:        484
Score:                 42.00          Matches:       6
Percent Similarity:    100.00%        Conservative:  0
Best Local Similarity: 100.00%        Mismatches:    0
Query Match:           100.00%        Indels:        0
DB:                    3              Gaps:          0

US-09-858-332G-28 (1-6) x BJ369193 (1-484)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         40 CACAATCACAACCACAAC 57

RESULT 503
BP036715
LOCUS       BP036715                 484 bp    mRNA    linear   EST 19-AUG-2004
DEFINITION  BP036715 Lotus corniculatus var. japonicus flower bud Lotus
            corniculatus var. japonicus cDNA clone MFB037e09_f 3', mRNA
            sequence.
ACCESSION   BP036715
VERSION     BP036715.1  GI:45413875
KEYWORDS    EST.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
```

rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae; Lotus.
REFERENCE 1 (bases 1 to 484)
AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
TITLE Characteristics of the Lotus japonicus gene repertoire deduced from large-scale expressed sequence tag (EST) analysis
JOURNAL Plant Mol. Biol. 54 (3), 405-414 (2004)
PUBMED 15284495
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES Location/Qualifiers
source 1. .484
/organism="Lotus corniculatus var. japonicus"
/mol_type="mRNA"
/isolate="Miyakojima MG-20"
/db_xref="taxon:34305"
/clone="MFB037e09_f"
/tissue_type="flower bud"
/clone_lib="Lotus corniculatus var. japonicus flower bud"
ORIGIN

Alignment Scores:

Pred. No.:	2.8e+03	Length:	484
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	3	Gaps:	0

US-09-858-332G-28 (1-6) x BP036715 (1-484)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        290 CATAATCATAATCATAAT 307
```

RESULT 504
BU010485
LOCUS BU010485 484 bp mRNA linear EST 22-AUG-2002
DEFINITION QGJ13J11.yg.ab1 QG_EFGHJ lettuce serriola Lactuca sativa cDNA clone QGJ13J11, mRNA sequence.
ACCESSION BU010485
VERSION BU010485.1 GI:22444880
KEYWORDS EST.
SOURCE Lactuca sativa
ORGANISM Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; campanulids; Asterales; Asteraceae; Cichorioideae; Cichorieae; Lactuca.
REFERENCE 1 (bases 1 to 484)
AUTHORS Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L., Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J., Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z., Church,S., Jackson,L. and Bradford,K.
TITLE Lettuce and Sunflower ESTs from the Compositae Genome Project http://compgenomics.ucdavis.edu/
JOURNAL Unpublished (2002)
COMMENT Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
belongs to contig QG_CA_Contig3547, see http://cgpdb.ucdavis.edu/ for details.
Plate: QGJ13 row: J column: 11.
FEATURES Location/Qualifiers
source 1. .484
/organism="Lactuca sativa"
/mol_type="mRNA"
/cultivar="L.serriola"
/db_xref="taxon:4236"
/clone="QGJ13J11"
/lab_host="E.coli"
/clone_lib="QG_EFGHJ lettuce serriola"
/note="Vector: pBRcDNASfiAB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at http://cgpdb.ucdavis.edu/ TAG_SEQ=Not found"
ORIGIN

Alignment Scores:

Pred. No.:	2.8e+03	Length:	484
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	5	Gaps:	0

US-09-858-332G-28 (1-6) x BU010485 (1-484)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        123 CACAATCACAATCACAAC 140
```

RESULT 505
AA949708
LOCUS AA949708 485 bp mRNA linear EST 19-APR-2001
DEFINITION LD29690.5prime LD Drosophila melanogaster embryo pOT2 Drosophila melanogaster cDNA clone LD29690 5prime similar to X63731: D.melanogaster regulatory signal sequence (clone scs), mRNA sequence.
ACCESSION AA949708
VERSION AA949708.1 GI:3112317
KEYWORDS EST.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 485)
AUTHORS Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G., Lewis,S. and Rubin,G.M.
TITLE BDGP/HHMI Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 296 row: H column: 6
High quality sequence stop: 484.
FEATURES Location/Qualifiers
source 1. .485
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="LD29690"
/sex="male and female"
/dev_stage="0 to 24 hours mixed stage embryonic"
/lab_host="XL1 Blue"
/clone_lib="LD Drosophila melanogaster embryo pOT2"
/note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2. "

```
ORIGIN

Alignment Scores:
Pred. No.:              2.8e+03       Length:        485
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     1             Gaps:          0

US-09-858-332G-28 (1-6) x AA949708 (1-485)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        191 CACAACCATAATCACAAT 208

RESULT 506
AA259786
LOCUS       AA259786                486 bp    mRNA    linear   EST 18-MAR-1997
DEFINITION  va87c12.r1 Soares mouse 3NME12 5 Mus musculus cDNA clone
            IMAGE:746422 5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0 KD
            PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION   AA259786
VERSION     AA259786.1  GI:1896272
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 486)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:455406
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 475.
FEATURES             Location/Qualifiers
     source          1. .486
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:746422"
                     /sex="unknown"
                     /tissue_type="fetus"
                     /dev_stage="12.5dpc total fetus"
                     /lab_host="DH10B"
                     /clone_lib="Soares mouse 3NME12 5"
                     /note="Organ: whole fetus; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Not I - oligo(dT) primer
                     [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCCTTATTTTTTTTTTTTTTTTT
                     3'],  on total  mouse RNA [provided by Minoru Ko, Wayne
                     State Univ.]; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo.  "
ORIGIN

Alignment Scores:
Pred. No.:              2.81e+03      Length:        486
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     1             Gaps:          0

US-09-858-332G-28 (1-6) x AA259786 (1-486)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        108 CACAACCATAACCACAAC 125

RESULT 507
BJ386030/c
LOCUS       BJ386030                486 bp    mRNA    linear   EST 08-MAR-2002
DEFINITION  BJ386030 Dictyostelium discoideum cDNA library, CF Dictyostelium
            discoideum cDNA clone ddc56n21 3', mRNA sequence.
ACCESSION   BJ386030
VERSION     BJ386030.1  GI:19295414
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 486)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the culmination
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .486
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddc56n21"
                     /sex="mat A"
                     /dev_stage="Culmination stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, CF"
ORIGIN

Alignment Scores:
Pred. No.:              2.81e+03      Length:        486
Score:                  42.00         Matches:       6
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     3             Gaps:          0

US-09-858-332G-28 (1-6) x BJ386030 (1-486)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        455 CACAATCACAACCACAAC 438

RESULT 508
CF961117
LOCUS       CF961117                486 bp    mRNA    linear   EST 15-JAN-2004
DEFINITION  11158rsiced_2453.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF961117
VERSION     CF961117.1  GI:38476984
KEYWORDS    EST.
```

```
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 486)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
  PUBMED    15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 486
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .486
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="trefoil"
                     /clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA
                     library"
ORIGIN

Alignment Scores:
Pred. No.:                2.81e+03        Length:         486
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       7               Gaps:           0

US-09-858-332G-28 (1-6) x CF961117 (1-486)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        178 CACAACCACAACCACAAC 195

RESULT 509
CX186077/c
LOCUS       CX186077                 486 bp    mRNA    linear   EST 28-DEC-2004
DEFINITION  D02_45-26_08.ab1 leaf inoculated with Marssonia pathogen of Populus
            euramericana Populus euramericana cDNA, mRNA sequence.
ACCESSION   CX186077
VERSION     CX186077.1  GI:56833501
KEYWORDS    EST.
SOURCE      Populus euramericana
  ORGANISM  Populus euramericana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Malpighiales; Salicaceae; Saliceae; Populus.
REFERENCE   1  (bases 1 to 486)
  AUTHORS   Huang,M., Zhang,X., Zhuge,Q., Zhou,Y., Zhou,H., Zhang,L., Guan,Y.,
            Zhang,B. and Wang,M.
  TITLE     Identification and Expression Analysis of EST-based Genes in the
            Leaf of poplar
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Minren Huang
            Key Laboratory Tree Genetic Engineering of Nanjing Forestry
            University
            Nanjing Forestry University
            Longpan road 9#, Nanjing, Jiangsu, P.R.China
            Tel: 086-025-85427412
            Fax: 086-025-85427412
            Email: mrhuang@njfu.edu.cn
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .486
                     /organism="Populus euramericana"
                     /mol_type="mRNA"
                     /db_xref="taxon:106131"
                     /clone_lib="leaf inoculated with Marssonia pathogen of
                     Populus euramericana"
                     /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
ORIGIN

Alignment Scores:
Pred. No.:                2.81e+03        Length:         486
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       8               Gaps:           0

US-09-858-332G-28 (1-6) x CX186077 (1-486)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        122 CACAACCACAACCACAAC 105

RESULT 510
CNS026FK/c
LOCUS       CNS026FK                 487 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            240J20 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL183305
VERSION     AL183305.1  GI:7821409
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  PUBMED    10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  PUBMED    10899143
REFERENCE   3  (bases 1 to 487)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
```

```
     source          1. .487
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="240J20"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG240DE10LP1
                     end : T7"
ORIGIN

Alignment Scores:
Pred. No.:              2.81e+03       Length:         487
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     10             Gaps:           0

US-09-858-332G-28 (1-6) x CNS026FK (1-487)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        136 CACAACCACAACCACAAC 119

RESULT 511
BJ441676/c
LOCUS       BJ441676                488 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ441676 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv47a13 3', mRNA sequence.
ACCESSION   BJ441676
VERSION     BJ441676.1  GI:19416398
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 488)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax:. 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .488
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv47a13"
                     /sex="mat A"
                     /dev_stage="Growth phase"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:              2.82e+03       Length:         488
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     3              Gaps:           0

US-09-858-332G-28 (1-6) x BJ441676 (1-488)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        464 CACAATCACAACCACAAC 447

RESULT 512
AZ321550
LOCUS       AZ321550                488 bp    DNA     linear   GSS 29-SEP-2000
DEFINITION  1M0042I11F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0042I11 F, genomic survey sequence.
ACCESSION   AZ321550
VERSION     AZ321550.1  GI:10374406
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 488)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
            Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
            Niederhausern,A. and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0042  row: I  column: 11
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 488.
FEATURES             Location/Qualifiers
     source          1. .488
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0042I11"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
ORIGIN

Alignment Scores:
Pred. No.:              2.82e+03       Length:         488
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     9              Gaps:           0
```

```
US-09-858-332G-28 (1-6) x AZ321550 (1-488)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         55 CACAACCACAACCACAAC 72

RESULT 513
AA044548
LOCUS       AA044548                489 bp    mRNA    linear   EST 05-SEP-1996
DEFINITION  mj12c02.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:475874 5' similar to gb:M38188 OVARIAN GRANULOSA CELL
            13.0 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION   AA044548
VERSION     AA044548.1  GI:1522788
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 489)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:286618
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 488.
FEATURES             Location/Qualifiers
     source          1. .489
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:475874"
                     /sex="unknown"
                     /tissue_type="embryo"
                     /dev_stage="13.5-14.5dpc total fetus"
                     /lab_host="DH10B"
                     /clone_lib="Soares mouse embryo NbME13.5 14.5"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTTT
                     T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                     14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                     State Univ., from 2 ]; double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector. Library went through one round of
                     normalization, and was constructed by Bento Soares and
                     M.Fatima Bonaldo.  "
ORIGIN

Alignment Scores:
Pred. No.:              2.83e+03        Length:         489
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AA044548 (1-489)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        136 CACAACCATAACCACAAC 153

RESULT 514
BX865324/c
LOCUS       BX865324                489 bp    mRNA    linear   EST 24-MAY-2004
DEFINITION  BX865324 tcba Oncorhynchus mykiss cDNA clone tcba0024c.m.01 5prim,
            mRNA sequence.
ACCESSION   BX865324
VERSION     BX865324.2  GI:42774785
KEYWORDS    EST.
SOURCE      Oncorhynchus mykiss (rainbow trout)
  ORGANISM  Oncorhynchus mykiss
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
REFERENCE   1  (bases 1 to 489)
  AUTHORS   Govoroun,M., Guiguen,Y. and Le Gac,F.
  TITLE     Construction and primary characterization of normalized cDNA
            libraries in rainbow trout, Oncorhynchus mykiss
  JOURNAL   Unpublished (2003)
COMMENT     On Dec 16, 2003 this sequence version replaced gi:39964638.
            Contact: Guiguen Y
            INRA - SCRIBE
            Campus de beaulieu, RENNES cedex, 35042, France
            Tel: 02.23.48.50.09
            Fax: 02.23.48.50.20
            Email: Yann.Guiguen@beaulieu.rennes.inra.fr
            Sequence cleaned of vector, adaptator and repetitions. Contact us
            at sigenasupport@jouy.inra.fr to obtain the chromatogram of this
            sequence.
            Plate: 0024  row: m  column: 1
            Seq primer: M13R.
FEATURES             Location/Qualifiers
     source          1. .489
                     /organism="Oncorhynchus mykiss"
                     /mol_type="mRNA"
                     /db_xref="taxon:8022"
                     /clone="tcba0024c.m.01"
                     /tissue_type="multi-tissues"
                     /dev_stage="from embryos to adults"
                     /lab_host="DH10B"
                     /clone_lib="tcba"
                     /note="Vector: pT7T3D-pac; Rainbow trout multi-tissues -
                     normalized + 1 subtraction; Clone distribution : AGENAE
                     Resource centre. Francois PIUMI,
                     Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et
                     Etude du genome (LREG), Domaine de Vilvert, 78352,
                     Jouy-en-Josas cedex, FRANCE, +33 (0) 1.34.65.28.02, +33
                     (0) 1.34.65.22.73"
ORIGIN

Alignment Scores:
Pred. No.:              2.83e+03        Length:         489
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     5               Gaps:           0

US-09-858-332G-28 (1-6) x BX865324 (1-489)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        362 CACAACCATAATCACAAT 345

RESULT 515
DR399696/c
LOCUS       DR399696                489 bp    mRNA    linear   EST 27-JUN-2005
```

```
DEFINITION  TKN105B10_636376 Taraxacum kok-saghyz root cDNA library Taraxacum
            kok-saghyz cDNA clone TKN105B10 5, mRNA sequence.
ACCESSION   DR399696
VERSION     DR399696.1  GI:68258940
KEYWORDS    EST.
SOURCE      Taraxacum kok-saghyz
  ORGANISM  Taraxacum kok-saghyz
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
            Cichorieae; Taraxacum.
REFERENCE   1  (bases 1 to 489)
  AUTHORS   Shintani,D.
  TITLE     Using EST from Taraxacum kok-saghyz root cDNA library to generate
            candidate rubber biosynthetic genes
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Shintani D
            Shintani Lab
            University of Nevada, Reno
            Dept. of Biochemistry/ms200, Reno, Nevada 89557, USA
            Tel: -775-784-4631
            Fax: -775-784-1650
            Email: shintani@unr.edu
            PCR PRimers
            FORWARD: pTriplex2 5' primer BD Biosciences
            BACKWARD: pTriplex2 3' primer BD Biosciences
            Plate: 105  row: B  column: 10
            Seq primer: BD Smart IV
            High quality sequence stop: 489.
FEATURES             Location/Qualifiers
     source          1. .489
                     /organism="Taraxacum kok-saghyz"
                     /mol_type="mRNA"
                     /db_xref="taxon:333970"
                     /clone="TKN105B10"
                     /tissue_type="root"
                     /dev_stage="1.2 to 3 month"
                     /clone_lib="Taraxacum kok-saghyz root cDNA library"
                     /note="Vector: pGADT7Rec-AB; Site_1: SfiIA; Site_2: SfiIB;
                     Total RNA was used for ds cDNA synthesis using Clontech
                     SMART protocol (Zhu et. al. (2001) Biotechniques 30:
                     892-897). SMART prepared amplified cDNA was then
                     normalized using DSN normalization method (Zhulidov et.
                     al., (2004) Nucleic Acid Res. 32(3): e37. Normalization
                     uncluded cDNA denaturation/reassociation, treatment of
                     duplex specific nuclease (DSN) (Shagin et. al. (2002)
                     Genome Res. 12: 1935-1942) and amplicification of
                     normalized fraction by PCR. Resulted normalized cDNA was
                     digested with SfiI and subcloned into pGADT7Rec-AB."
ORIGIN

Alignment Scores:
Pred. No.:              2.83e+03        Length:         489
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     8               Gaps:           0

US-09-858-332G-28 (1-6) x DR399696 (1-489)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        293 CATAACCATAACCATAAC 276

RESULT 516
AG966239
LOCUS       AG966239                489 bp    DNA     linear   GSS 01-FEB-2005
DEFINITION  Drosophila simulans DNA, clone: DSM1-002B05.R.fa, genomic survey
            sequence.
ACCESSION   AG966239
VERSION     AG966239.1  GI:58513820
KEYWORDS    GSS.
SOURCE      Drosophila simulans
  ORGANISM  Drosophila simulans
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1
  AUTHORS   Hattori,M., Toyoda,A., Murakami,K., Kuroki,Y., Fujiyama,A.,
            Toshio,T.K. and Sakaki,Y.
  TITLE     BAC end sequences of Library DNB1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 489)
  AUTHORS   Hattori,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JAN-2005) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the BAC library DSM1
            For BAC library availability, please contact Masa-Toshi Yamamoto
            (yamamoto@kit.jp).
            Submitted (30-11-2004) by  Masahira Hattori,
            RIKEN, Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp,Tel:81-45-503-9111,
            Fax:81-45-503-9170)
            This work was done in collaboration with Yamamoto, M-T. Drosophila
            Genetic Resource Center
            Saga Ippongi-cho, Ukyo-ku, Kyoto 616-8354, Japan
            Tel: 81-75-873-2660 FAX: 81-75-861-0881
            PRIMERS
                 Sequencing : R
            LIBRARY
                 Vector     : pKS150
                 R.Site 1   : SacI
                 R.Site 2   : SacI.
FEATURES             Location/Qualifiers
     source          1. .489
                     /organism="Drosophila simulans"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7240"
                     /clone="DSM1-002B05.R.fa"
                     /clone_lib="DSM1 Drosophila BAC library"
ORIGIN

Alignment Scores:
Pred. No.:              2.83e+03        Length:         489
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     10              Gaps:           0

US-09-858-332G-28 (1-6) x AG966239 (1-489)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        231 CATAACCATAACCATAAT 248

RESULT 517
AV557891/c
LOCUS       AV557891                491 bp    mRNA    linear   EST 23-FEB-2004
DEFINITION  AV557891 Arabidopsis thaliana green siliques Columbia Arabidopsis
            thaliana cDNA clone SQ083e03F 3', mRNA sequence.
ACCESSION   AV557891
VERSION     AV557891.1  GI:8729306
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 491)
```

```
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     A large scale analysis of cDNA in Arabidopsis thaliana: Generation
            of 12,028 non-redundant expressed sequence tags from normalized and
            size-selected cDNA libraries
  JOURNAL   DNA Res. 7 (3), 175-180 (2000)
   PUBMED   10907847
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .491
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /ecotype="Columbia"
                     /db_xref="taxon:3702"
                     /clone="SQ083e03F"
                     /tissue_type="green siliques"
                     /clone_lib="Arabidopsis thaliana green siliques Columbia"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
ORIGIN

Alignment Scores:
Pred. No.:                2.84e+03        Length:         491
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       1               Gaps:           0

US-09-858-332G-28 (1-6) x AV557891 (1-491)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        309 CACAACCACAACCACAAC 292

RESULT 518
BP043778
LOCUS       BP043778                491 bp    mRNA    linear   EST 19-AUG-2004
DEFINITION  BP043778 Lotus corniculatus var. japonicus flower bud Lotus
            corniculatus var. japonicus cDNA clone MFBL049f07_f 3', mRNA
            sequence.
ACCESSION   BP043778
VERSION     BP043778.1  GI:45577672
KEYWORDS    EST.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1  (bases 1 to 491)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Characteristics of the Lotus japonicus gene repertoire deduced from
            large-scale expressed sequence tag (EST) analysis
  JOURNAL   Plant Mol. Biol. 54 (3), 405-414 (2004)
   PUBMED   15284495
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .491
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="mRNA"
                     /isolate="Miyakojima MG-20"
                     /db_xref="taxon:34305"
                     /clone="MFBL049f07_f"
                     /tissue_type="flower bud"
                     /clone_lib="Lotus corniculatus var. japonicus flower bud"
ORIGIN

Alignment Scores:
Pred. No.:                2.84e+03        Length:         491
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       3               Gaps:           0

US-09-858-332G-28 (1-6) x BP043778 (1-491)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         42 CACAATCATAATCATAAT 59

RESULT 519
CF963249
LOCUS       CF963249                491 bp    mRNA    linear   EST 15-JAN-2004
DEFINITION  12143rsiced_6341.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF963249
VERSION     CF963249.1  GI:38479116
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 491)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 491
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .491
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="trefoil"
                     /clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA
                     library"
ORIGIN

Alignment Scores:
Pred. No.:                2.84e+03        Length:         491
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       7               Gaps:           0

US-09-858-332G-28 (1-6) x CF963249 (1-491)
```

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        172 CACAACCACAACCACAAC 189

RESULT 520
BF292932
LOCUS       BF292932                492 bp    mRNA    linear   EST 17-NOV-2000
DEFINITION  WHE2166_E09_J18ZS Triticum turgidum L. var. durum (durum wheat)
            whole plant cDNA library Triticum turgidum cDNA clone
            WHE2166_E09_J18, mRNA sequence.
ACCESSION   BF292932
VERSION     BF292932.1  GI:11223996
KEYWORDS    EST.
SOURCE      Triticum turgidum
  ORGANISM  Triticum turgidum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 492)
  AUTHORS   Akhunov,E., Anderson,O.D., Chao,S., Chin,A., Choi,D.W., Close,T.J.,
            Fenton,R.D., Han,P.S., Hsia,C.C., Kang,Y., Kianian,P., Lazo,G.R.,
            Miller,R., Otto,C., Rausch,C.J., Seaton,C.L., Simons,K., Tong,J.C.
            and Zhang,D.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Whole plant cDNA library from Triticum turgidum L. var.
            durum
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene SK primer.
FEATURES             Location/Qualifiers
     source          1. .492
                     /organism="Triticum turgidum"
                     /mol_type="mRNA"
                     /cultivar="Langdon-16"
                     /db_xref="taxon:4571"
                     /clone="WHE2166_E09_J18"
                     /tissue_type="All tissues"
                     /dev_stage="Different growth stages"
                     /lab_host="E. coli SOLR"
                     /clone_lib="Triticum turgidum L. var. durum (durum wheat)
                     whole plant cDNA library"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; Plants were grown in a growth
                     chamber at North Dakota State University  (Kianian, Otto,
                     Simons).  Tissues collected from seven-day etiolated
                     seedling leaf, stem, root and seed; leaf from plant at
                     fourth leaf  stage; spike from pre-anthesis through 20
                     days after anthesis; flag  leaf; leaf and stem tissue from
                     tillers. and root.  Total RNA and  poly(A) RNA were
                     prepared from each tissue and then pooled, a cDNA  library
                     was made, and the cDNA clones were in vivo excised to give
                     pBluescript phagemids in the TJ Close lab (Akhunov, Chin,
                     Choi, Close,  Fenton, Kianian, Otto, Simons, Zhang) at the
                     University of California,  Riverside.  Plasmid DNA
                     preparations and DNA sequencing were performed  in the OD
                     Anderson lab (all other authors)."
ORIGIN

Alignment Scores:
Pred. No.:              2.84e+03        Length:         492
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0
```

```
US-09-858-332G-28 (1-6) x BF292932 (1-492)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        269 CACAACCACAACCACAAT 286

RESULT 521
BJ358288/c
LOCUS       BJ358288                492 bp    mRNA    linear   EST 07-MAR-2002
DEFINITION  BJ358288 Dictyostelium discoideum cDNA library, AF Dictyostelium
            discoideum cDNA clone dda56k20 3', mRNA sequence.
ACCESSION   BJ358288
VERSION     BJ358288.1  GI:19257883
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 492)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the aggregation
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .492
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dda56k20"
                     /sex="mat A"
                     /dev_stage="Aggregation stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN

Alignment Scores:
Pred. No.:              2.84e+03        Length:         492
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0
```

```
US-09-858-332G-28 (1-6) x BJ358288 (1-492)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        274 CACAATCACAACCACAAC 257

RESULT 522
BJ435953/c
LOCUS       BJ435953                492 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ435953 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv29j03 3', mRNA sequence.
ACCESSION   BJ435953
VERSION     BJ435953.1  GI:19410675
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 492)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
```

```
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES              Location/Qualifiers
     source           1. .492
                      /organism="Dictyostelium discoideum"
                      /mol_type="mRNA"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /clone="ddv29j03"
                      /sex="mat A"
                      /dev_stage="Growth phase"
                      /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:              2.84e+03        Length:         492
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-858-332G-28 (1-6) x BJ435953 (1-492)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        387 CACAATCACAATCACAAT 370

RESULT 523
CB868647
LOCUS       CB868647                492 bp    mRNA    linear   EST 03-JUL-2003
DEFINITION  HC08P11w CH Hordeum vulgare cDNA clone HC08P11 5-PRIME, mRNA
            sequence.
ACCESSION   CB868647
VERSION     CB868647.1  GI:30070624
KEYWORDS    EST.
SOURCE      Hordeum vulgare
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 492)
  AUTHORS   Varshney,R.K., Zhang,H., Burton,R., Stein,N., Langridge,P. and
            Graner,A.
  TITLE     Barley ESTs from coleoptile tissue
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Stein Nils
            Molecular Markers Group, Department Genbank
            Institute of Plant Genetics and Crop Plant Research (IPK)
            Correnssstr. 3, 06466, Gatersleben, Germany
            Tel: 039482-5522
            Fax: 039482-5595
            Email: stein@ipk-gatersleben.de
            Insert Length: 492   Std Error: 0.00
            Plate: 8  row: P  column: 11
            Seq primer: T7.
FEATURES              Location/Qualifiers
     source           1. .492
                      /organism="Hordeum vulgare"
                      /mol_type="mRNA"
                      /cultivar="Sloop"
                      /db_xref="GABI:547049"
                      /db_xref="taxon:4513"
                      /clone="HC08P11"
                      /tissue_type="coleoptile"
                      /dev_stage="coleoptile, 3 days old"
                      /lab_host="DH10B"
                      /clone_lib="CH"
                      /note="Vector: pSPORT; Site_1: SalI (5-end of cDNA);
                      Site_2: NotI (3-end of cDNA); Due to the cloning system
                      used blue/white selection for recombinants is not 100%
                      reliable. Average insert size is 1.3 kb."
ORIGIN

Alignment Scores:
Pred. No.:              2.84e+03        Length:         492
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x CB868647 (1-492)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        272 CACAACCACAACCACAAC 289

RESULT 524
CB964758/c
LOCUS       CB964758                492 bp    mRNA    linear   EST 29-APR-2003
DEFINITION  NL11_D05 Drought stress (leaf) Oryza sativa (indica cultivar-group)
            cDNA clone NL11_D05 3', mRNA sequence.
ACCESSION   CB964758
VERSION     CB964758.1  GI:30226866
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 492)
  AUTHORS   Markandeya,G., Ravindra Babu,P., Venkat Reddy,B., Nagabhushana,I.,
            Chandra Sekhar,A., Bennetzen,J.L., Ramakrishna,W. and Reddy,A.R.
  TITLE     ESTs from a normalized cDNA library of drought stressed rice
            seedlings (Oryza sativa L.cv Nagina 22)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Reddy AR
            Department of Plant Sciences, School of Life Sciences
            University of Hyderabad
            P.O. Central University, Hyderabad-500 046, A.P, India
            Tel: 0091-40-3010265
            Fax: 0091-40-3010145
            Email: arjulsl@uohyd.ernet.in
            Insert Length: 492   Std Error: 0.00
            Seq primer: CGCCAGGGTTTTCCCAGTCACGAC.
FEATURES              Location/Qualifiers
     source           1. .492
                      /organism="Oryza sativa (indica cultivar-group)"
                      /mol_type="mRNA"
                      /cultivar="Nagina 22"
                      /db_xref="taxon:39946"
                      /clone="NL11_D05"
                      /tissue_type="Entire leaf tissue"
                      /dev_stage="35 day-old seedlings"
                      /clone_lib="Drought stress (leaf)"
                      /note="Organ: Leaf; Vector: T7T3Pac; ESTs from normalized
                      leaf cDNA Library from drought stressed seedlings "
ORIGIN

Alignment Scores:
Pred. No.:              2.84e+03        Length:         492
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x CB964758 (1-492)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
```

```
Db        361 CACAACCACAACCACAAC 344

RESULT 525
W46041
LOCUS       W46041                  492 bp    mRNA    linear   EST 23-MAY-1996
DEFINITION  mc82h02.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:355059 5' similar to gb:M38188 OVARIAN GRANULOSA CELL
            13.0 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION   W46041
VERSION     W46041.1  GI:1330757
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 492)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:226859
            Seq primer: BTPrimer
            High quality sequence stop: 489.
FEATURES             Location/Qualifiers
     source          1. .492
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:355059"
                     /sex="unknown"
                     /tissue_type="embryo"
                     /dev_stage="13.5-14.5dpc total fetus"
                     /lab_host="DH10B"
                     /clone_lib="Soares mouse embryo NbME13.5 14.5"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTTTT
                     T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                     14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                     State Univ., from 2 ]; double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector. Library went through one round of
                     normalization, and was constructed by Bento Soares and
                     M.Fatima Bonaldo.  "
ORIGIN

Alignment Scores:
Pred. No.:              2.84e+03        Length:         492
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     8               Gaps:           0

US-09-858-332G-28 (1-6) x W46041 (1-492)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        133 CACAACCATAACCACAAC 150

RESULT 526
BU885165
LOCUS       BU885165                493 bp    mRNA    linear   EST 17-OCT-2002
DEFINITION  R021E09 Populus root cDNA library Populus tremula x Populus
            tremuloides cDNA 5 prime, mRNA sequence.
ACCESSION   BU885165
VERSION     BU885165.1  GI:24076682
KEYWORDS    EST.
SOURCE      Populus tremula x Populus tremuloides
  ORGANISM  Populus tremula x Populus tremuloides
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Malpighiales; Salicaceae; Saliceae; Populus.
REFERENCE   1  (bases 1 to 493)
  AUTHORS   Unneberg,P., Bhalerao,R.R., Jansson,S. and Sterky,F.
  TITLE     The poplar tree transcriptome: Analysis of expressed sequence tags
            from multiple libraries
  JOURNAL   Unpublished (2002)
COMMENT     Contact: BHALERAO RUPALI R.
            Umea Plant Science Center
            Department of Plant Physiology
            University of Umea, 901 87 Umea, Sweden
            Tel: +46 90 786 5279
            Fax: +46 90 786 6676
            Email: rupali.bhalerao@plantphys.umu.se.
FEATURES             Location/Qualifiers
     source          1. .493
                     /organism="Populus tremula x Populus tremuloides"
                     /mol_type="mRNA"
                     /db_xref="taxon:47664"
                     /tissue_type="root"
                     /clone_lib="Populus root cDNA library"
ORIGIN

Alignment Scores:
Pred. No.:              2.85e+03        Length:         493
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     5               Gaps:           0

US-09-858-332G-28 (1-6) x BU885165 (1-493)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        139 CACAACCACAACCACAAC 156

RESULT 527
CB826688
LOCUS       CB826688                493 bp    mRNA    linear   EST 16-MAY-2003
DEFINITION  LjNEST65g4r Lotus japonicus nodule library 5 and 7 week-old Lotus
            corniculatus var. japonicus cDNA 5', mRNA sequence.
ACCESSION   CB826688
VERSION     CB826688.1  GI:29964928
KEYWORDS    EST.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1  (bases 1 to 493)
  AUTHORS   Colebatch,G., Freund,S., Trevaskis,B and Udvardi,M.
  TITLE     Lotus japonicus root nodule ESTs: tools for functional genomics
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Udvardi MK
            Molecular Plant Nutrition
            Max Planck Institute of Molecular Plant Physiology
            Am Muehlenberg 1, 14476 Golm, Germany
```

```
            Fax: 49 331 567 8250
            Email: udvardi@mpimp-golm.mpg.de
            Seq primer: T7
            High quality sequence stop: 493.
FEATURES             Location/Qualifiers
     source          1. .493
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="mRNA"
                     /cultivar="Gifu (B-129)"
                     /db_xref="taxon:34305"
                     /dev_stage="5 and 7 week-old plants"
                     /clone_lib="Lotus japonicus nodule library 5 and 7
                     week-old"
                     /note="Organ: Nodule; Vector: pSPORT1; Site_1: SalI;
                     Site_2: NotI; The library was prepared using mRNA
                     extracted from nodules of 5 and 7 week-old Lotus plants.
                     Nodules were induced by, and contained Mesorhizobium
                     strain R7A."
ORIGIN

Alignment Scores:
Pred. No.:                2.85e+03        Length:          493
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       6               Gaps:            0

US-09-858-332G-28 (1-6) x CB826688 (1-493)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        163 CATAATCATAACCATAAT 180

RESULT 528
BJ388347
LOCUS       BJ388347                494 bp    mRNA    linear   EST 08-MAR-2002
DEFINITION  BJ388347 Dictyostelium discoideum cDNA library, SF Dictyostelium
            discoideum cDNA clone dds9h04 5', mRNA sequence.
ACCESSION   BJ388347
VERSION     BJ388347.1  GI:19299433
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 494)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the slug stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .494
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dds9h04"
                     /sex="mat A"
                     /dev_stage="Slug stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, SF"
ORIGIN

Alignment Scores:
Pred. No.:                2.86e+03        Length:          494
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       3               Gaps:            0

US-09-858-332G-28 (1-6) x BJ388347 (1-494)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        400 CACAACCACAACCACAAC 417

RESULT 529
CO310824/c
LOCUS       CO310824                494 bp    mRNA    linear   EST 28-JUN-2004
DEFINITION  EK222720.5prime Exelixis FlyTag CK01 pCDNA-SK+ Drosophila
            melanogaster cDNA clone EK222720 5, mRNA sequence.
ACCESSION   CO310824
VERSION     CO310824.1  GI:49363618
KEYWORDS    EST.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 494)
  AUTHORS   Kopczynski,C., Platt,D., Campbell,J., Muzong,C., Laufer,A.,
            Peterson,E. and Swimmer,C.
  TITLE     Exelixis FlyTag EST Project CK01 Library
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: EK.2227  row: B  column: 8
            High quality sequence stop: 415.
FEATURES             Location/Qualifiers
     source          1. .494
                     /organism="Drosophila melanogaster"
                     /mol_type="mRNA"
                     /db_xref="taxon:7227"
                     /clone="EK222720"
                     /clone_lib="Exelixis FlyTag CK01 pCDNA-SK+"
                     /note="Organ: mixed stage embryos, imaginal disks, and
                     adult heads; Vector: pCDNA-SK+; Site_1: NotI; Site_2:
                     XhoI; Random primed, normalized library from mixed stage
                     embryos, imaginal disks, and adult heads."
ORIGIN

Alignment Scores:
Pred. No.:                2.86e+03        Length:          494
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       7               Gaps:            0

US-09-858-332G-28 (1-6) x CO310824 (1-494)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        484 CATAATCATAATCATAAT 467

RESULT 530
BM409788
LOCUS       BM409788                495 bp    mRNA    linear   EST 22-JAN-2002
DEFINITION  EST584115 tomato breaker fruit Lycopersicon esculentum cDNA clone
            cLEG50I17 5' end, mRNA sequence.
ACCESSION   BM409788
VERSION     BM409788.1  GI:18261418
KEYWORDS    EST.
SOURCE      Lycopersicon esculentum (Solanum lycopersicum)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
```

```
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 495)
  AUTHORS   Alcala,J., Vrebalov,J., White,R., Vision,T., Karamycheva,S.A.,
            Tsai,J., Bougri,O., Kirkness,E., Utterback,T., Van Aken,S.,
            Ronning,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and
            Giovannoni,J.
  TITLE     Generation of ESTs from tomato fruit tissue, breaker stage (2002)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html
            This clone is available through the Clemson University Genomics
            Institute
            Seq primer: T3.
FEATURES              Location/Qualifiers
     source           1. .495
                      /organism="Lycopersicon esculentum"
                      /mol_type="mRNA"
                      /cultivar="TA496"
                      /db_xref="taxon:4081"
                      /clone="cLEG50I17"
                      /tissue_type="Pericarp"
                      /dev_stage="breaker"
                      /lab_host="SOLR"
                      /clone_lib="tomato breaker fruit"
                      /note="Vector: pBluescriptSKmCUadapt; Site_1: EcoR1;
                      Site_2: XhoI; supplier: Boyce Thompson Institute;
                      sequencing: The Institute for Genomic Research.  Fruit
                      were harvested at the breaker stage (first sign of
                      lycopene accumulation on the blossom end of fruit).  Fruit
                      were cut in half and the seeds and locules were discarded
                      prior to freezing the pericarp."
ORIGIN

Alignment Scores:
Pred. No.:                2.86e+03        Length:         495
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       3               Gaps:           0

US-09-858-332G-28 (1-6) x BM409788 (1-495)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        422 CACAATCACAATCACAAT 439

RESULT 531
CF960382
LOCUS       CF960382                495 bp    mRNA    linear   EST 15-JAN-2004
DEFINITION  12469rsiced_7450.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF960382
VERSION     CF960382.1  GI:38476249
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 495)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
  PUBMED    15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 495
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .495
                      /organism="Oryza sativa (indica cultivar-group)"
                      /mol_type="mRNA"
                      /cultivar="LYP9"
                      /db_xref="taxon:39946"
                      /tissue_type="whole plant"
                      /dev_stage="trefoil"
                      /clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA
                      library"
ORIGIN

Alignment Scores:
Pred. No.:                2.86e+03        Length:         495
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       7               Gaps:           0

US-09-858-332G-28 (1-6) x CF960382 (1-495)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        132 CACAACCACAACCACAAC 149

RESULT 532
C90140
LOCUS       C90140                  496 bp    mRNA    linear   EST 20-APR-1998
DEFINITION  C90140 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
            discoideum cDNA clone SSI136, mRNA sequence.
ACCESSION   C90140
VERSION     C90140.1  GI:3059760
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 496)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
            Tel: 81-298-53-4664
            Fax: 81-298-53-6614
            Email: hideko@biol.tsukuba.ac.jp.
FEATURES              Location/Qualifiers
     source           1. .496
                      /organism="Dictyostelium discoideum"
                      /mol_type="mRNA"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /clone="SSI136"
                      /dev_stage="slug"
                      /clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
ORIGIN
```

```
Alignment Scores:
Pred. No.:                2.87e+03        Length:         496
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       5               Gaps:           0

US-09-858-332G-28 (1-6) x C90140 (1-496)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        201 CACAATCACAATCACAAT 218

RESULT 533
CD886243
LOCUS       CD886243                 496 bp    mRNA    linear   EST 14-JUL-2003
DEFINITION  G118.101L23F010601 G118 Triticum aestivum cDNA clone G118101L23,
            mRNA sequence.
ACCESSION   CD886243
VERSION     CD886243.1  GI:32651629
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 496)
  AUTHORS   Genoplante.
  TITLE     Genoplante, a major partnership french program in plant genomics
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Genoplante
            Genoplante
            93, rue Henri Rochefort 91025 EVRY CEDEX France
            Tel: 33 1 69 47 54 00
            Fax: 33 1 69 47 54 10
            This sequence has been generated in the framework of the french
            plant genomics programme 'Genoplante' (http://www.genoplante.com
            and http://genoplante-info.infobiogen.fr).
FEATURES             Location/Qualifiers
     source          1. .496
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /cultivar="recital"
                     /db_xref="taxon:4565"
                     /clone="G118101L23"
                     /tissue_type="grain (118 degrees per day after
                     pollination)"
                     /clone_lib="G118"
ORIGIN

Alignment Scores:
Pred. No.:                2.87e+03        Length:         496
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x CD886243 (1-496)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        247 CACAACCACAACCACAAC 264

RESULT 534
CF988679
LOCUS       CF988679                 497 bp    mRNA    linear   EST 04-FEB-2005
DEFINITION  17338rsicee_15193.y1 Oryza sativa cv. LYP9 tillering whole plant
            cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF988679
VERSION     CF988679.1  GI:58590371
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 497)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
  PUBMED    15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 497
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .497
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="tillering"
                     /clone_lib="Oryza sativa cv. LYP9 tillering whole plant
                     cDNA library"
ORIGIN

Alignment Scores:
Pred. No.:                2.88e+03        Length:         497
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       7               Gaps:           0

US-09-858-332G-28 (1-6) x CF988679 (1-497)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        187 CACAACCACAACCACAAC 204

RESULT 535
CO266833
LOCUS       CO266833                 497 bp    mRNA    linear   EST 23-JUN-2004
DEFINITION  EK147719.5prime Exelixis FlyTag CK01 pCDNA-SK+ Drosophila
            melanogaster cDNA clone EK147719 5, mRNA sequence.
ACCESSION   CO266833
VERSION     CO266833.1  GI:49152418
KEYWORDS    EST.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 497)
  AUTHORS   Kopczynski,C., Platt,D., Campbell,J., Muzong,C., Laufer,A.,
            Peterson,E. and Swimmer,C.
  TITLE     Exelixis FlyTag EST Project CK01 Library
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Stapleton, M.
```

```
           BDGP
           Lawrence Berkeley National Lab
           One Cyclotron Rd, Berkeley, CA 94720, USA
           Fax: 510 486 6798
           Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
           Plate: EK.1477  row: B  column: 7
           High quality sequence stop: 445.
FEATURES             Location/Qualifiers
     source          1. .497
                     /organism="Drosophila melanogaster"
                     /mol_type="mRNA"
                     /db_xref="taxon:7227"
                     /clone="EK147719"
                     /clone_lib="Exelixis FlyTag CK01 pCDNA-SK+"
                     /note="Organ: mixed stage embryos, imaginal disks, and
                     adult heads; Vector: pCDNA-SK+; Site_1: NotI; Site_2:
                     XhoI; Random primed, normalized library from mixed stage
                     embryos, imaginal disks, and adult heads."
ORIGIN

Alignment Scores:
Pred. No.:              2.88e+03        Length:         497
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     7               Gaps:           0

US-09-858-332G-28 (1-6) x CO266833 (1-497)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         34 CATAACCATAATCATAAC 51

RESULT 536
AQ184606
LOCUS       AQ184606                497 bp    DNA     linear   GSS 01-NOV-1998
DEFINITION  HS_2212_A1_A11_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2212 Col=21 Row=A, genomic survey
            sequence.
ACCESSION   AQ184606
VERSION     AQ184606.1  GI:3584048
KEYWORDS    GSS..
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 497)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 2212  row: A  column: 21
            Class: BAC ends
            High quality sequence stop: 497.
FEATURES             Location/Qualifiers
     source          1. .497
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=2212 Col=21 Row=A"
                     /sex="male"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
ORIGIN

Alignment Scores:
Pred. No.:              2.88e+03        Length:         497
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     9               Gaps:           0

US-09-858-332G-28 (1-6) x AQ184606 (1-497)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        156 CACAACCATAATCACAAT 173

RESULT 537
AW979496
LOCUS       AW979496                498 bp    mRNA    linear   EST 18-MAY-2001
DEFINITION  EST310517 tomato root deficiency, Cornell University Lycopersicon
            esculentum cDNA clone cLEW2I13 5', mRNA sequence.
ACCESSION   AW979496
VERSION     AW979496.1  GI:8171015
KEYWORDS    EST.
SOURCE      Lycopersicon esculentum (Solanum lycopersicum)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 498)
  AUTHORS   van der Hoeven,R.S., Garvin,D.F., Matern,A.L., Holt,I.E., Liang,F.,
            Upton,J., Hansen,T.S., Ronning,C.M., Craven,M.B., Bowman,C.L.,
            Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Giovannoni,J.J.
            and Tanksley,S.D.
  TITLE     Generation of ESTs from tomato nutrient-deficient roots
  JOURNAL   Unpublished (1999)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html
            5 prime sequence.
FEATURES             Location/Qualifiers
     source          1. .498
                     /organism="Lycopersicon esculentum"
                     /mol_type="mRNA"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cLEW2I13"
                     /tissue_type="roots"
                     /dev_stage="5-6 weeks old"
                     /clone_lib="tomato root deficiency, Cornell University"
                     /note="Vector: pBluescript SK-; Site_1: 5' EcoRI; Site_2:
                     3' XhoI; supplier: Tanksley; Tissue supplied by Dave
                     Garvin (USDA-ARS, Ithaca, NY 14850). Roots were harvested
                     from plants grown under the following
                     deficiencies/stresses: 10 mM Al, Zn, P, K, Fe, N), and
                     mRNA was isolated from indivdual treatments. Proportional
                     aliquots of mRNA of each treatment were mixed and used for
                     library construction."
ORIGIN

Alignment Scores:
Pred. No.:              2.88e+03        Length:         498
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0
```

```
US-09-858-332G-28 (1-6) x AW979496 (1-498)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         40 CACAATCACAATCACAAT 57

RESULT 538
BU014670/c
LOCUS       BU014670                498 bp    mRNA    linear   EST 22-AUG-2002
DEFINITION  QGJ7P15.yg.ab1 QG_EFGHJ lettuce serriola Lactuca sativa cDNA clone
            QGJ7P15, mRNA sequence.
ACCESSION   BU014670
VERSION     BU014670.1  GI:22449065
KEYWORDS    EST.
SOURCE      Lactuca sativa
  ORGANISM  Lactuca sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
            Cichorieae; Lactuca.
REFERENCE   1  (bases 1 to 498)
  AUTHORS   Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J.,
            Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y.,
            Lai,Z., Church,S., Jackson,L. and Bradford,K.
  TITLE     Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
            belongs to contig QG_CA_Contig1494, see http://cgpdb.ucdavis.edu/
            for details.
            Plate: QGJ7  row: P  column: 15.
FEATURES             Location/Qualifiers
     source          1. .498
                     /organism="Lactuca sativa"
                     /mol_type="mRNA"
                     /cultivar="L.serriola"
                     /db_xref="taxon:4236"
                     /clone="QGJ7P15"
                     /lab_host="E.coli"
                     /clone_lib="QG_EFGHJ lettuce serriola"
                     /note="Vector: pBRcDNASfiAB; The library was constructed
                     from 10 different sources of RNA from a single genotype.
                     Separate cDNAs were generated using primers that
                     incorporated unique 5' and 3' tags to distinguish each
                     source of RNA. cDNAs were then pooled, size-fractionated,
                     directionally cloned into a custom medium-copy vector and
                     transformations made with four size classes to minimize
                     size bias. Details of each source of RNA and library
                     construction can be obtained at http://cgpdb.ucdavis.edu/
                     TAG_TISSUE=flowers pre-fertilized
                     TAG_LIB=QG_EFGHJ lettuce serriola
                     TAG_SEQ=GCTTGACGGG"
ORIGIN

Alignment Scores:
Pred. No.:                2.88e+03        Length:          498
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       5               Gaps:            0

US-09-858-332G-28 (1-6) x BU014670 (1-498)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        289 CACAATCACAATCACAAT 272

RESULT 539
BX099249
LOCUS       BX099249                498 bp    mRNA    linear   EST 06-FEB-2003
DEFINITION  BX099249 Soares infant brain 1NIB Homo sapiens cDNA clone
            IMAGp998A19147 ; IMAGE:29528, mRNA sequence.
ACCESSION   BX099249
VERSION     BX099249.1  GI:27829885
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 498)
  AUTHORS   Ebert,L., Heil,O., Hennig,S., Neubert,P., Partsch,E., Peters,M.,
            Radelof,U., Schneider,D. and Korn,B.
  TITLE     Human UnigeneSet - RZPD3
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Ina Rolfs
            RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
            Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
            RZPD; IMAGp998A19147.
            RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
            Human UnigeneSet - RZPD3 (RZPDLIB No.972)
            http://www.rzpd.de/CloneCards/cgi-
            bin/showLib.pl.cgi/response?libNo=972 Contact: Ina Rolfs
            RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
            Heubnerweg 6, D-14059 Berlin, Germany
            Tel: +49 30 32639 101
            Fax: +49 30 32639 111
            www.rzpd.de
            This clone is available royalty-free from RZPD;
            contact RZPD (clone@rzpd.de) for further information. Seq primer:
            M13u, Primer sequence: CGTTGTAAAACGACGGCCAGT.
FEATURES             Location/Qualifiers
     source          1. .498
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGp998A19147 ; IMAGE:29528"
                     /sex="female"
                     /dev_stage="73 days post natal"
                     /lab_host="DH10B (ampicillin resistant)"
                     /clone_lib="Soares infant brain 1NIB"
                     /note="Organ: whole brain; Vector: Lafmid BA; Site_1: Not
                     I; Site_2: Hind III; 1st strand cDNA was primed with a Not
                     I - oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTT 3'];
                     double-stranded cDNA was ligated to Hind III adaptors
                     (Pharmacia), digested with Not I and directionally cloned
                     into the Not I and Hind III sites of the Lafmid BA vector.
                     Library went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
ORIGIN

Alignment Scores:
Pred. No.:                2.88e+03        Length:          498
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       5               Gaps:            0

US-09-858-332G-28 (1-6) x BX099249 (1-498)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        205 CATAATCATAATCATAAT 222

RESULT 540
```

```
CR491970
LOCUS       CR491970                498 bp    DNA     linear   GSS 17-NOV-2004
DEFINITION  mth2-165J11RM1 BAC end, cultivar Jemalong A17 of Medicago
            truncatula, genomic survey sequence.
ACCESSION   CR491970
VERSION     CR491970.1  GI:48653546
KEYWORDS    GSS.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 498)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUN-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
FEATURES             Location/Qualifiers
     source          1. .498
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /cultivar="Jemalong A17"
                     /db_xref="taxon:3880"
                     /clone_lib="MTH2"
                     /note="Vector: pBeloBAC11 ; Site_1: HindIII ; Site_2:
                     HindIII ; Cook, D.R. and Kim, D.J
                     Genoscope sequence ID : mth2-165J11RM1"
ORIGIN

Alignment Scores:
Pred. No.:                2.88e+03        Length:         498
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       11              Gaps:           0

US-09-858-332G-28 (1-6) x CR491970 (1-498)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         80 CATAATCATAATCATAAT 97

RESULT 541
BQ489630
LOCUS       BQ489630                499 bp    mRNA    linear   EST 07-JUN-2002
DEFINITION  21-E9233-006-008-J05-T3 Sugar beet MPIZ-ADIS-006 Lambda Zap II
            library Beta vulgaris cDNA clone J-5-8, mRNA sequence.
ACCESSION   BQ489630
VERSION     BQ489630.1  GI:21334250
KEYWORDS    EST.
SOURCE      Beta vulgaris
  ORGANISM  Beta vulgaris
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            Caryophyllales; Amaranthaceae; Beta.
REFERENCE   1  (bases 1 to 499)
  AUTHORS   Bellin,D., Werber,M., Theis,T., Weisshaar,B. and Schneider,K.
  TITLE     EST sequencing, annotation and macroarray expression analysis of
            more than 3000 sugar beet cDNAs identifies genes with root-specific
            expression pattern
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Weisshaar B
            ADIS DNA core facility at MPIZ
            Max-Planck-Institute for Plant Breeding Research
            Carl-von-Linne Weg 10, 50829 Koeln, Germany
            Fax: 00492215062851
            Email: weisshaa@mpiz-koeln.mpg.de
            Seq primer: T3 'AATTAACCCTCACTAAAGGG'
            High quality sequence stop: 499.
FEATURES             Location/Qualifiers
     source          1. .499
                     /organism="Beta vulgaris"
                     /mol_type="mRNA"
                     /db_xref="taxon:161934"
                     /clone="J-5-8"
                     /dev_stage="4 week old pot-grown plants"
                     /clone_lib="Sugar beet MPIZ-ADIS-006 Lambda Zap II
                     library"
                     /note="Organ: shoot and root; Vector: pBluescript SK- from
                     lambda ZAP II; cDNA (lambda ZAP-II) library from sugar
                     beet, whole plant mRNA, Prepared using the Stratagene
                     UniZAP cDNA kit, cloning sites EcoRI-XhoI, primer sites
                     and orientation:
                     rev-T3-SacI-SK-EcoRI-GGCACGAGG-5pr-cDNA-polyA-XhoI-KpnI-T7
                     -uni"
ORIGIN

Alignment Scores:
Pred. No.:                2.89e+03        Length:         499
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       5               Gaps:           0

US-09-858-332G-28 (1-6) x BQ489630 (1-499)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         98 CACAATCACAATCACAAT 115

RESULT 542
BU991516/c
LOCUS       BU991516                499 bp    mRNA    linear   EST 22-OCT-2002
DEFINITION  HD07C14r HD Hordeum vulgare cDNA clone HD07C14 5-PRIME, mRNA
            sequence.
ACCESSION   BU991516
VERSION     BU991516.1  GI:24242462
KEYWORDS    EST.
SOURCE      Hordeum vulgare
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 499)
  AUTHORS   Zhang,H., Weschke,W., Michalek,W., Stein,N. and Graner,A.
  TITLE     EST sequencing and analysis in barley (2002)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Stein Nils
            Molecular Markers Group, Department Genbank
            Institute of Plant Genetics and Crop Plant Research (IPK)
            Corrensstr. 3, 06466, Gatersleben, Germany
            Tel: 039482-5522
            Fax: 039482-5595
            Email: stein@ipk-gatersleben.de
            Insert Length: 499   Std Error: 0.00
            Plate: 7  row: C  column: 14
            Seq primer: M13rev.
FEATURES             Location/Qualifiers
     source          1. .499
                     /organism="Hordeum vulgare"
                     /mol_type="mRNA"
                     /cultivar="Golden Promise"
                     /db_xref="GABI:241623"
                     /db_xref="taxon:4513"
                     /clone="HD07C14"
                     /tissue_type="callus"
                     /dev_stage="callus (5-10 mm in diameter)"
                     /lab_host="XL10-Gold"
                     /clone_lib="HD"
                     /note="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of
                     cDNA); Site_2: XhoI (3'-end of cDNA); Due to a cloning
                     artefact caused by the kit, in most cases the EcoRI site
```

is NOT present, as well as the EcoRIadapter used for cloning. To excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI,PstI). NOTE: Also due to the cloning system used Blue/white selection for recombinats is not 100% reliable.Average insert size is 1 kb"
ORIGIN

Alignment Scores:
Pred. No.: 2.89e+03 Length: 499
Score: 42.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 5 Gaps: 0

US-09-858-332G-28 (1-6) x BU991516 (1-499)

Qy 1 HisAsnHisAsnHisAsn 6
||||||||||||||||||
Db 334 CATAATCATAATCATAAC 317

RESULT 543
CF991700
LOCUS CF991700 499 bp mRNA linear EST 04-FEB-2005
DEFINITION 16342rsicee_13610.y1 Oryza sativa cv. LYP9 tillering whole plant cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA sequence.
ACCESSION CF991700
VERSION CF991700.1 GI:58593392
KEYWORDS EST.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 499)
AUTHORS Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W., Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X., Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J., Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q., Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z., Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P., Wu,S. and Liu,J.
TITLE The Genomes of Oryza sativa: A History of Duplications
JOURNAL PLoS Biol. 3 (2), e38 (2005)
PUBMED 15685292
COMMENT Contact: Yan Zhou
Bioinformatics Department
Hangzhou Genomics Institute
No.51 Zhijiang Road, Hangzhou 310008, China
Tel: 86-571-56805886
Fax: 86-571-56805884
Email: zhouyan@genomics.org.cn
Seq primer: M13 Forward
High quality sequence stop: 499
POLYA=No.
FEATURES Location/Qualifiers
source 1. .499
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/cultivar="LYP9"
/db_xref="taxon:39946"
/tissue_type="whole plant"
/dev_stage="tillering"
/clone_lib="Oryza sativa cv. LYP9 tillering whole plant cDNA library"
ORIGIN

Alignment Scores:
Pred. No.: 2.89e+03 Length: 499
Score: 42.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0

US-09-858-332G-28 (1-6) x CF991700 (1-499)

Qy 1 HisAsnHisAsnHisAsn 6
||||||||||||||||||
Db 207 CACAACCACAACCACAAC 224

RESULT 544
CV548502
LOCUS CV548502 499 bp mRNA linear EST 22-OCT-2004
DEFINITION 3032HAEMEX55F6 BgORESTES infected NHM 3032 haemocytes Biomphalaria glabrata cDNA clone ZBA6487, mRNA sequence.
ACCESSION CV548502
VERSION CV548502.1 GI:54425098
KEYWORDS EST.
SOURCE Biomphalaria glabrata
ORGANISM Biomphalaria glabrata
Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Basommatophora; Lymnaeoidea; Planorbidae; Biomphalaria.
REFERENCE 1 (bases 1 to 499)
AUTHORS Lockyer,A.E., Spinks,J.N., Kane,R.A., Dias Neto,E., Noble,L.R., Rollinson,D. and Jones,C.S.
TITLE ESTs from Biomphalaria glabrata using the ORESTES approach
JOURNAL Unpublished (2003)
COMMENT Contact: Lockyer, A.E.
Wolfson Wellcome Biomedical Laboratory
The Natural History Museum
Cromwell Road, London, SW7 5BD, UK
Tel: +44 (0)20 7942 5148
Fax: +44 (0)20 7942 5518
Email: a.lockyer@nhm.ac.uk
High quality sequence stop: 499
POLYA=No.
FEATURES Location/Qualifiers
source 1. .499
/organism="Biomphalaria glabrata"
/mol_type="mRNA"
/strain="NHM 3032"
/db_xref="taxon:6526"
/clone="ZBA6487"
/sex="Hermaphrodite"
/tissue_type="haemocytes"
/clone_lib="BgORESTES infected NHM 3032 haemocytes"
/note="Vector: pGEM; mRNA, extracted from infected 3032 (susceptible) snails was used as a template for RT-PCR with random primers to generate cDNA fragments. These were cloned and sequenced using M13F. Primer sequences were removed from sequence."
ORIGIN

Alignment Scores:
Pred. No.: 2.89e+03 Length: 499
Score: 42.00 Matches: 6
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0

US-09-858-332G-28 (1-6) x CV548502 (1-499)

Qy 1 HisAsnHisAsnHisAsn 6
||||||||||||||||||
Db 460 CACAACCACAACCACAAC 477

RESULT 545
CC175193/c
LOCUS CC175193 499 bp DNA linear GSS 02-MAY-2003
DEFINITION ZMMBBc0292L05r ZMMBBc Zea mays genomic clone ZMMBBc0292L05 3', genomic survey sequence.
ACCESSION CC175193

```
VERSION     CC175193.1  GI:30274953
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 499)
  AUTHORS   Bharti,A.K., Young,S., Kavchok,S., Keizer,G., Bronzino,A.C.,
            Sugiyama,R., Fuks,G., Yu,Y., Wing,R. and Messing,J.
  TITLE     Sequencing of the maize genome at PGIR (2003a)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Bharti,A.K.
            Dr.Joachim Messing's lab
            The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers
            University
            190 Frelinghuysen Road, Piscataway, NJ 08854, USA
            Tel: 732 445 3801
            Fax: 732 445 5735
            Email: bharti@waksman.rutgers.edu
            Seq primer: SP6
            Class: BAC ends
            High quality sequence start: 103.
FEATURES             Location/Qualifiers
     source          1. .499
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBBc0292L05"
                     /lab_host="E. coli DH10B"
                     /clone_lib="ZMMBBc"
                     /note="Vector: pTARBAC1.3; Site_1: BamHI; Site_2: BamHI"
ORIGIN

Alignment Scores:
Pred. No.:              2.89e+03        Length:         499
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     9               Gaps:           0

US-09-858-332G-28 (1-6) x CC175193 (1-499)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        328 CACAACCACAACCACAAT 311

RESULT 546
AA110311
LOCUS       AA110311                500 bp    mRNA    linear   EST 02-OCT-1997
DEFINITION  mp51d09.r1 Barstead MPLRB1 Mus musculus cDNA clone IMAGE:572753 5'
            similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0 KD PROTEIN HGR74
            (HUMAN);, mRNA sequence.
ACCESSION   AA110311
VERSION     AA110311.1  GI:1662053
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI: 347401
            Seq primer: -28m13 rev2 ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .500
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="BALB/c"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:572753"
                     /sex="mixed"
                     /tissue_type="Kidney"
                     /dev_stage="6 weeks"
                     /lab_host="DH10B"
                     /clone_lib="Barstead MPLRB1"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     [CATGGATTCGGTACC], digested with Not I and cloned into the
                     Not I and Eco RI sites of the modified pT7T3 vector.
                     Library constructed by Bob Barstead."
ORIGIN

Alignment Scores:
Pred. No.:              2.89e+03        Length:         500
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AA110311 (1-500)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        249 CACAACCATAACCACAAC 266

RESULT 547
BJ365115
LOCUS       BJ365115                500 bp    mRNA    linear   EST 08-MAR-2002
DEFINITION  BJ365115 Dictyostelium discoideum cDNA library, CF Dictyostelium
            discoideum cDNA clone ddc34o02 5', mRNA sequence.
ACCESSION   BJ365115
VERSION     BJ365115.1  GI:19274415
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the culmination
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .500
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
```

```
                     /db_xref="taxon:44689"
                     /clone="ddc34o02"
                     /sex="mat A"
                     /dev_stage="Culmination stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, CF"
ORIGIN

Alignment Scores:
Pred. No.:                2.89e+03        Length:         500
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       3               Gaps:           0

US-09-858-332G-28 (1-6) x BJ365115 (1-500)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        439 CACAACCACAACCACAAC 456

RESULT 548
BP188517/c
LOCUS       BP188517                 500 bp    mRNA    linear   EST 17-JUL-2003
DEFINITION  BP188517 planarian head cDNA Dugesia japonica cDNA clone 03878_HH,
            mRNA sequence.
ACCESSION   BP188517
VERSION     BP188517.1  GI:32903041
KEYWORDS    EST.
SOURCE      Dugesia japonica
  ORGANISM  Dugesia japonica
            Eukaryota; Metazoa; Platyhelminthes; Turbellaria; Seriata;
            Tricladida; Paludicola; Dugesiidae; Dugesia.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   Mineta,K., Nakazawa,M., Cebria,F., Ikeo,K., Agata,K. and
            Gojobori,T.
  TITLE     Origin and evolutionary process of the CNS elucidated by
            comparative genomics analysis of planarian ESTs
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 100 (13), 7666-7671 (2003)
   PUBMED   12802012
COMMENT     Contact: Katsuhiko Mineta
            National Institute of Genetics, Center for Information Biology and
            DNA Data Bank of Japan
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6847
            Fax: 81-559-81-6848
            Email: kmineta@lab.nig.ac.jp
            These clones and additional information are obtained from our web
            site: http://www.cib.nig.ac.jp/dda/.
FEATURES             Location/Qualifiers
     source          1. .500
                     /organism="Dugesia japonica"
                     /mol_type="mRNA"
                     /db_xref="taxon:6161"
                     /clone="03878_HH"
                     /tissue_type="head"
                     /dev_stage="adult"
                     /clone_lib="planarian head cDNA"
ORIGIN

Alignment Scores:
Pred. No.:                2.89e+03        Length:         500
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       3               Gaps:           0

US-09-858-332G-28 (1-6) x BP188517 (1-500)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        339 CATAATCATAATCATAAC 322

RESULT 549
BU991888/c
LOCUS       BU991888                 500 bp    mRNA    linear   EST 22-OCT-2002
DEFINITION  HD08D17r HD Hordeum vulgare cDNA clone HD08D17 5-PRIME, mRNA
            sequence.
ACCESSION   BU991888
VERSION     BU991888.1  GI:24242834
KEYWORDS    EST.
SOURCE      Hordeum vulgare
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   Zhang,H., Weschke,W., Michalek,W., Stein,N. and Graner,A.
  TITLE     EST sequencing and analysis in barley (2002)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Stein Nils
            Molecular Markers Group, Department Genbank
            Institute of Plant Genetics and Crop Plant Research (IPK)
            Corrensstr. 3, 06466, Gatersleben, Germany
            Tel: 039482-5522
            Fax: 039482-5595
            Email: stein@ipk-gatersleben.de
            Insert Length: 500   Std Error: 0.00
            Plate: 8  row: D  column: 17
            Seq primer: M13rev.
FEATURES             Location/Qualifiers
     source          1. .500
                     /organism="Hordeum vulgare"
                     /mol_type="mRNA"
                     /cultivar="Golden Promise"
                     /db_xref="GABI:241996"
                     /db_xref="taxon:4513"
                     /clone="HD08D17"
                     /tissue_type="callus"
                     /dev_stage="callus (5-10 mm in diameter)"
                     /lab_host="XL10-Gold"
                     /clone_lib="HD"
                     /note="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of
                     cDNA); Site_2: XhoI (3'-end of cDNA); Due to a cloning
                     artefact caused by the kit, in most cases the EcoRI site
                     is NOT present, as well as the EcoRIadapter used for
                     cloning. To excise the insert, restriction sites upstream
                     EcoRI should be used (e.g. BamHI, SalI,PstI). NOTE: Also
                     due to the cloning system used Blue/white selection for
                     recombinats is not 100% reliable.Average insert size is 1
                     kb"
ORIGIN

Alignment Scores:
Pred. No.:                2.89e+03        Length:         500
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       5               Gaps:           0

US-09-858-332G-28 (1-6) x BU991888 (1-500)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        334 CATAATCATAATCATAAC 317

RESULT 550
CB660142/c
LOCUS       CB660142                 500 bp    mRNA    linear   EST 09-APR-2003
DEFINITION  OSJNEd01K10.f OSJNEd Oryza sativa (japonica cultivar-group) cDNA
            clone OSJNEd01K10 5', mRNA sequence.
ACCESSION   CB660142
VERSION     CB660142.1  GI:29663867
KEYWORDS    EST.
```

```
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   Jantasuriyarat,C., Gowda,M., Haller,K., Hatfield,J., Lu,G.,
            Stahlberg,E., Zhou,B., Li,H., Kim,H., Yu,Y., Dean,R.A., Wing,R.A.,
            Soderlund,C. and Wang,G.L.
  TITLE     Large-scale identification of expressed sequence tags involved in
            rice and rice blast fungus interaction
  JOURNAL   Plant Physiol. 138 (1), 105-115 (2005)
   PUBMED   15888683
COMMENT     Contact: Rod Wing
            Arizona Genomics Institute
            University of Arizona
            Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
            85721-0088, USA
            Tel: 520 626 3967
            Fax: 520 621 9288
            Email: rwing@genome.arizona.edu
            PCR PRimers
            FORWARD: gta aaa cga cgg cca gtg
            BACKWARD: gga aac agc tat gac cat g
            Plate: 01  row: K  column: 10
            Seq primer: gta aaa cga cgg cca gtg.
FEATURES             Location/Qualifiers
     source          1. .500
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="OSJNEd01K10"
                     /tissue_type="Leaf"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
                     /clone_lib="OSJNEd"
                     /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
                     XhoI; 24 hrs after innoculation with Rice Blast (C9240-1)"
ORIGIN

Alignment Scores:
Pred. No.:                2.89e+03        Length:         500
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x CB660142 (1-500)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         74 CACAACCACAACCACAAC 57

RESULT 551
CK041708
LOCUS       CK041708                500 bp    mRNA    linear   EST 05-FEB-2005
DEFINITION  48024rsiceh_17347.y1 Oryza sativa cv. PA64s panicle ferile cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CK041708
VERSION     CK041708.1  GI:58653028
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 500
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .500
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="PA64s"
                     /db_xref="taxon:39946"
                     /tissue_type="panicle"
                     /cell_type="ferile"
                     /dev_stage="heading/flowering"
                     /clone_lib="Oryza sativa cv. PA64s panicle ferile cDNA
                     library"
ORIGIN

Alignment Scores:
Pred. No.:                2.89e+03        Length:         500
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       7               Gaps:           0

US-09-858-332G-28 (1-6) x CK041708 (1-500)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        217 CACAACCACAACCACAAC 234

RESULT 552
CL892616/c
LOCUS       CL892616                500 bp    DNA     linear   GSS 30-AUG-2004
DEFINITION  abg16c02.y1 Soybean random, unfiltered genomic library Glycine max
            genomic, genomic survey sequence.
ACCESSION   CL892616
VERSION     CL892616.1  GI:51639657
KEYWORDS    GSS.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   Nunberg,A., Bedell,J.A., Citek,R.W., Robbins,D., McMenamy,J.,
            Peterson,S., Jones,J., Fries,J., Budiman,M.A., Nguyen,H. and
            Stacey,G.
  TITLE     Methylation filtered genomic sequences from Glycine max
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Gary Stacey
            University of Missouri
            108 Waters Hall, Columbia, MO 65211, USA
            Tel: 573-884-1267
            Fax: 573-882-0588
            Email: staceyg@missouri.edu
            LidID: 230
            Class: shotgun.
FEATURES             Location/Qualifiers
```

```
     source          1. .500
                     /organism="Glycine max"
                     /mol_type="genomic DNA"
                     /cultivar="Williams 82"
                     /db_xref="taxon:3847"
                     /tissue_type="Young leaves"
                     /clone_lib="Soybean random, unfiltered genomic library"
                     /note="Vector: pOT2; Site_1: BstXI; Randomly sheared
                     genomic DNA ranging from 0.7-1.5 kb were end repaired and
                     ligated to BstXI linkers prior to cloning in BstXI-cut
                     pOT2. LibID: 230"
ORIGIN

Alignment Scores:
Pred. No.:                2.89e+03        Length:         500
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       10              Gaps:           0

US-09-858-332G-28 (1-6) x CL892616 (1-500)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        161 CACAACCACAACCACAAT 144

RESULT 553
CD888217
LOCUS       CD888217                501 bp    mRNA    linear   EST 14-JUL-2003
DEFINITION  G118.107G09F010606 G118 Triticum aestivum cDNA clone G118107G09,
            mRNA sequence.
ACCESSION   CD888217
VERSION     CD888217.1  GI:32655318
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 501)
  AUTHORS   Genoplante.
  TITLE     Genoplante, a major partnership french program in plant genomics
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Genoplante
            Genoplante
            93, rue Henri Rochefort 91025 EVRY CEDEX France
            Tel: 33 1 69 47 54 00
            Fax: 33 1 69 47 54 10
            This sequence has been generated in the framework of the french
            plant genomics programme 'Genoplante' (http://www.genoplante.com
            and http://genoplante-info.infobiogen.fr).
FEATURES              Location/Qualifiers
     source           1. .501
                      /organism="Triticum aestivum"
                      /mol_type="mRNA"
                      /cultivar="recital"
                      /db_xref="taxon:4565"
                      /clone="G118107G09"
                      /tissue_type="grain (118 degrees per day after
                      pollination)"
                      /clone_lib="G118"
ORIGIN

Alignment Scores:
Pred. No.:                2.9e+03         Length:         501
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-858-332G-28 (1-6) x CD888217 (1-501)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        334 CACAACCACAACCACAAT 351

RESULT 554
BJ435042/c
LOCUS       BJ435042                502 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ435042 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv25j24 3', mRNA sequence.
ACCESSION   BJ435042
VERSION     BJ435042.1  GI:19409764
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 502)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES              Location/Qualifiers
     source           1. .502
                      /organism="Dictyostelium discoideum"
                      /mol_type="mRNA"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /clone="ddv25j24"
                      /sex="mat A"
                      /dev_stage="Growth phase"
                      /clone_lib="Dictyostelium discoideum cDNA library, VF"
ORIGIN

Alignment Scores:
Pred. No.:                2.91e+03        Length:         502
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       3               Gaps:           0

US-09-858-332G-28 (1-6) x BJ435042 (1-502)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        301 CACAATCACAATCACAAT 284

RESULT 555
CJ346215
LOCUS       CJ346215                502 bp    mRNA    linear   EST 15-JUN-2005
DEFINITION  CJ346215 Molgula tectiformis unpublished cDNA library, cleaving
            embryo Molgula tectiformis cDNA clone mtcl010114 5', mRNA sequence.
ACCESSION   CJ346215
VERSION     CJ346215.1  GI:67786363
KEYWORDS    EST.
SOURCE      Molgula tectiformis
  ORGANISM  Molgula tectiformis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
            Stolidobranchia; Molgulidae; Molgula.
REFERENCE   1  (bases 1 to 502)
  AUTHORS   Gyoja,F., Satou,Y. and Satoh,N.
  TITLE     Expressed genes in Molgula tectiformis
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
```

```
           1111 Yata, Mishima, Shizuoka 411-8540, Japan
           Tel: 81-559-81-6856
           Fax: 81-559-81-6855
           Email: tshini@genes.nig.ac.jp
           When you want to obtain this EST clone, please send an e-mail to
           Nori Satoh (satoh@ascidian.zool.kyoto-u.ac.jp) and its cc to Yutaka
           Satou (yutaka@ascidian.zool.kyoto-u.ac.jp).
FEATURES             Location/Qualifiers
     source          1. .502
                     /organism="Molgula tectiformis"
                     /mol_type="mRNA"
                     /db_xref="taxon:30286"
                     /clone="mtcl010l14"
                     /tissue_type="whole animal"
                     /dev_stage="cleaving embryo"
                     /clone_lib="Molgula tectiformis unpublished cDNA library,
                     cleaving embryo"
ORIGIN

Alignment Scores:
Pred. No.:              2.91e+03       Length:         502
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     7              Gaps:           0

US-09-858-332G-28 (1-6) x CJ346215 (1-502)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        374 CACAATCACAATCACAAC 391

RESULT 556
CW260551/c
LOCUS       CW260551                502 bp    DNA     linear   GSS 30-OCT-2004
DEFINITION  104_727_11229028_116_35196_010 Sorghum methylation filtered library
            (LibID: 104) Sorghum bicolor genomic clone 11229028, genomic survey
            sequence.
ACCESSION   CW260551
VERSION     CW260551.1  GI:54969378
KEYWORDS    GSS.
SOURCE      Sorghum bicolor (sorghum)
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 502)
  AUTHORS   Bedell,J.A., Budiman,M.A., Nunberg,A., Citek,R.W., Robbins,D.,
            Jones,J., Flick,E., Rohlfing,T., Fries,J., Bradford,K.,
            McMenamy,J., Smith,M., Holeman,H., Roe,B.A, Wiley,G., Korf,I.F.,
            Rabinowicz,P.D., Lakey,N., McCombie,W.R., Jeddeloh,J.A. and
            Martienssen,R.A.
  TITLE     Sorghum genome sequencing by methylation filtration
  JOURNAL   PLoS Biol. 3 (1), e13 (2005)
   PUBMED   15660154
COMMENT     Contact: Bedell JA
            Orion Genomics, LLC
            4041 Forest Park Ave, St. Louis, MO 63108, USA
            Tel: 314 615 6979
            Fax: 314 615 5975
            Email: jbedell@oriongenomics.com
            Plate: 727  row: g  column: 04
            Seq primer: T3 Reverse
            Class: methylation filtered
            High quality sequence stop: 502.
FEATURES             Location/Qualifiers
     source          1. .502
                     /organism="Sorghum bicolor"
                     /mol_type="genomic DNA"
                     /cultivar="ATx623"
                     /db_xref="taxon:4558"
                     /clone="11229028"
                     /clone_lib="Sorghum methylation filtered library (LibID:
                     104)"
                     /note="Organ: leaf; Vector: pBCSK(-); Site_1: HincII; DNA
                     prepared from purified nuclei was randomly sheared,
                     end-repaired, size fractionated to enrich for the 0.5 to 5
                     kb fraction, ligated into HincII-digested pBCSK(-) vector
                     and electroporated into E. coli cells. This is a
                     methylation filtered library."
ORIGIN

Alignment Scores:
Pred. No.:              2.91e+03       Length:         502
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     10             Gaps:           0

US-09-858-332G-28 (1-6) x CW260551 (1-502)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        113 CATAACCATAACCATAAC 96

RESULT 557
CW339495/c
LOCUS       CW339495                502 bp    DNA     linear   GSS 31-OCT-2004
DEFINITION  104_840_11485302_148_36140_027 Sorghum methylation filtered library
            (LibID: 104) Sorghum bicolor genomic clone 11485302, genomic survey
            sequence.
ACCESSION   CW339495
VERSION     CW339495.1  GI:55055683
KEYWORDS    GSS.
SOURCE      Sorghum bicolor (sorghum)
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 502)
  AUTHORS   Bedell,J.A., Budiman,M.A., Nunberg,A., Citek,R.W., Robbins,D.,
            Jones,J., Flick,E., Rohlfing,T., Fries,J., Bradford,K.,
            McMenamy,J., Smith,M., Holeman,H., Roe,B.A, Wiley,G., Korf,I.F.,
            Rabinowicz,P.D., Lakey,N., McCombie,W.R., Jeddeloh,J.A. and
            Martienssen,R.A.
  TITLE     Sorghum genome sequencing by methylation filtration
  JOURNAL   PLoS Biol. 3 (1), e13 (2005)
   PUBMED   15660154
COMMENT     Contact: Bedell JA
            Orion Genomics, LLC
            4041 Forest Park Ave, St. Louis, MO 63108, USA
            Tel: 314 615 6979
            Fax: 314 615 5975
            Email: jbedell@oriongenomics.com
            Plate: 840  row: e  column: 06
            Seq primer: SWfor Forward
            Class: methylation filtered
            High quality sequence stop: 502.
FEATURES             Location/Qualifiers
     source          1. .502
                     /organism="Sorghum bicolor"
                     /mol_type="genomic DNA"
                     /cultivar="ATx623"
                     /db_xref="taxon:4558"
                     /clone="11485302"
                     /clone_lib="Sorghum methylation filtered library (LibID:
                     104)"
                     /note="Organ: leaf; Vector: pBCSK(-); Site_1: HincII; DNA
                     prepared from purified nuclei was randomly sheared,
                     end-repaired, size fractionated to enrich for the 0.5 to 5
                     kb fraction, ligated into HincII-digested pBCSK(-) vector
                     and electroporated into E. coli cells. This is a
                     methylation filtered library."
ORIGIN
```

```
Alignment Scores:
Pred. No.:                2.91e+03        Length:         502
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       10              Gaps:           0

US-09-858-332G-28 (1-6) x CW339495 (1-502)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         26 CACAACCACAACCACAAC 9

RESULT 558
AW908751
LOCUS       AW908751                503 bp    mRNA    linear   EST 25-MAY-2000
DEFINITION  uf57a05.y1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
            IMAGE:1515440 5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0
            KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION   AW908751
VERSION     AW908751.1  GI:8073984
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 503)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Other_ESTs: uf57a05.x1
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:942292
            Seq primer: -40RP from Gibco
            High quality sequence stop: 468.
FEATURES             Location/Qualifiers
     source          1. .503
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1515440"
                     /sex="female (lactating)"
                     /tissue_type="mammary gland"
                     /lab_host="DH10B"
                     /clone_lib="Soares_mammary_gland_NMLMG"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from mammary
                     gland tissue from a lactating female, and was then primed
                     with a Not I - oligo(dT) primer. Double-stranded  cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I  and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector.  Library is normalized.  Library
                     was constructed by Bento Soares and M. Fatima Bonaldo. "
ORIGIN

Alignment Scores:
Pred. No.:                2.91e+03        Length:         503
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       1               Gaps:           0

US-09-858-332G-28 (1-6) x AW908751 (1-503)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        117 CACAACCATAACCACAAC 134

RESULT 559
BQ106494
LOCUS       BQ106494                503 bp    mRNA    linear   EST 16-APR-2002
DEFINITION  fc1077.e Rose Petals (Fragrant Cloud) Lambda Zap Express Library
            Rosa hybrid cultivar cDNA clone fc1077.e 5', mRNA sequence.
ACCESSION   BQ106494
VERSION     BQ106494.1  GI:20156156
KEYWORDS    EST.
SOURCE      Rosa hybrid cultivar
  ORGANISM  Rosa hybrid cultivar
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Rosales; Rosaceae; Rosoideae; Rosa.
REFERENCE   1  (bases 1 to 503)
  AUTHORS   Guterman,I., Shalit,M., Menda,N., Piestun,D., Dafny-Yelin,M.,
            Shalev,G., Bar,E., Davydov,O., Ovadis,M., Emanuel,M., Wang,J.,
            Adam,Z., Pichersky,E., Lewinsohn,E., Zamir,D., Vainstein,A. and
            Weiss,D.
  TITLE     Rose Scent: Genomics Approach to Discovering Novel Floral
            Fragrance-Related Genes
  JOURNAL   Plant Cell 14 (10), 2325-2338 (2002)
COMMENT     Contact: Naama Menda
            Petal Genomics
            Faculty of Agricultural, Food and Environmental Quality Sciences,
            The Hebrew University of Jerusalem
            P.O. Box 12, Rehovot, 76100, Israel
            Tel: 972 8 9489 683
            Fax: 972 8 9468 265
            Email: shaham@agri.huji.ac.il
            Seq primer: T3 forward.
FEATURES             Location/Qualifiers
     source          1. .503
                     /organism="Rosa hybrid cultivar"
                     /mol_type="mRNA"
                     /strain="Fragrant Cloud"
                     /db_xref="taxon:128735"
                     /clone="fc1077.e"
                     /tissue_type="Petals"
                     /dev_stage="Young open flower at stage four"
                     /clone_lib="Rose Petals (Fragrant Cloud) Lambda Zap
                     Express Library"
                     /note="Vector: pBKCMV; Site_1: EcoRI; Site_2: XhoI"
ORIGIN

Alignment Scores:
Pred. No.:                2.91e+03        Length:         503
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       3               Gaps:           0

US-09-858-332G-28 (1-6) x BQ106494 (1-503)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        371 CACAACCACAACCACAAC 388

RESULT 560
BZ417164
LOCUS       BZ417164                503 bp    DNA     linear   GSS 10-DEC-2002
DEFINITION  if75g09.g2 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
            if75g09 5', genomic survey sequence.
ACCESSION   BZ417164
VERSION     BZ417164.1  GI:26360492
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
```

clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 503)
AUTHORS Rabinowicz,P.D., O'Shaughnessy,A.L., Balija,V., Dedhia,N., Katzenburger,F., King,L., Miller,B., Muller,S., Nascimento,L., Zutavern,T., McCombie,W.R. and Martienssen,R.A.
TITLE Genomic shotgun sequences from Zea mays (methyl-filtered)
JOURNAL Unpublished (2002)
COMMENT Contact: W. Richard McCombie
Lita Annenberg Hazen Genome Sequencing Center
Cold Spring Harbor Laboratory
PO Box 100, Cold Spring Harbor, NY 11724, USA
Tel: 516 367 8884
Fax: 516 367 8874
Email: mccombie@cshl.org
Plate: if75 row: g column: 09
Seq primer: -21M13UnivRev
Class: shotgun
High quality sequence stop: 503.
FEATURES Location/Qualifiers
source 1. .503
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="B73"
/db_xref="taxon:4577"
/clone="if75g09"
/lab_host="DH5a"
/clone_lib="WGS-ZmaysF (DH5a methyl filtered)"
/note="Organ: immature ears; Site_1: Xba I; Site_2: Xba I; The vector was digested with XbaI and one nucleotide was added by fill in in the recessive 3' end. The genomic DNA was nebulized, end repaired, adaptor ligated and size fractionated using sephadex. The resulting fragments were between 0.8 and 3 kb and were cloned into the vector (.x/y reads in M13mp19, .b/g reads in pUC19). The same ligation was transformed into DH5a."
ORIGIN

Alignment Scores:

Pred. No.:	2.91e+03	Length:	503
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

US-09-858-332G-28 (1-6) x BZ417164 (1-503)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        317 CATAACCATAACCATAAC 334
```

RESULT 561
BF293488
LOCUS BF293488 504 bp mRNA linear EST 17-NOV-2000
DEFINITION WHE2157_D05_H09ZS Triticum turgidum L. var. durum (durum wheat) whole plant cDNA library Triticum turgidum cDNA clone WHE2157_D05_H09, mRNA sequence.
ACCESSION BF293488
VERSION BF293488.1 GI:11224552
KEYWORDS EST.
SOURCE Triticum turgidum
ORGANISM Triticum turgidum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 504)
AUTHORS Akhunov,E., Anderson,O.D., Chao,S., Chin,A., Choi,D.W., Close,T.J., Fenton,R.D., Han,P.S., Hsia,C.C., Kang,Y., Kianian,P., Lazo,G.R., Miller,R., Otto,C., Rausch,C.J., Seaton,C.L., Simons,K., Tong,J.C. and Zhang,D.
TITLE The structure and function of the expressed portion of the wheat genomes - Whole plant cDNA library from Triticum turgidum L. var. durum
JOURNAL Unpublished (2000)
COMMENT Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oandersn@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
Seq primer: Stratagene SK primer.
FEATURES Location/Qualifiers
source 1. .504
/organism="Triticum turgidum"
/mol_type="mRNA"
/cultivar="Langdon-16"
/db_xref="taxon:4571"
/clone="WHE2157_D05_H09"
/tissue_type="All tissues"
/dev_stage="Different growth stages"
/lab_host="E. coli SOLR"
/clone_lib="Triticum turgidum L. var. durum (durum wheat) whole plant cDNA library"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site_1: EcoRI; Site_2: XhoI; Plants were grown in a growth chamber at North Dakota State University (Kianian, Otto, Simons). Tissues collected from seven-day etiolated seedling leaf, stem, root and seed; leaf from plant at fourth leaf stage; spike from pre-anthesis through 20 days after anthesis; flag leaf; leaf and stem tissue from tillers. and root. Total RNA and poly(A) RNA were prepared from each tissue and then pooled, a cDNA library was made, and the cDNA clones were in vivo excised to give pBluescript phagemids in the TJ Close lab (Akhunov, Chin, Choi, Close, Fenton, Kianian, Otto, Simons, Zhang) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."
ORIGIN

Alignment Scores:

Pred. No.:	2.92e+03	Length:	504
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-28 (1-6) x BF293488 (1-504)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        304 CACAACCACAACCACAAT 321
```

RESULT 562
BJ365128
LOCUS BJ365128 504 bp mRNA linear EST 08-MAR-2002
DEFINITION BJ365128 Dictyostelium discoideum cDNA library, CF Dictyostelium discoideum cDNA clone ddc34a09 5', mRNA sequence.
ACCESSION BJ365128
VERSION BJ365128.1 GI:19274428
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 504)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the culmination stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics

```
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES              Location/Qualifiers
     source           1. .504
                      /organism="Dictyostelium discoideum"
                      /mol_type="mRNA"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /clone="ddc34a09"
                      /sex="mat A"
                      /dev_stage="Culmination stage"
                      /clone_lib="Dictyostelium discoideum cDNA library, CF"
ORIGIN

Alignment Scores:
Pred. No.:                2.92e+03        Length:         504
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       3               Gaps:           0

US-09-858-332G-28 (1-6) x BJ365128 (1-504)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        470 CACAACCACAACCACAAC 487

RESULT 563
BQ253175
LOCUS       BQ253175                 504 bp    mRNA    linear   EST 05-JUL-2004
DEFINITION  san70d03.y1 Gm-c1052 Glycine max cDNA clone SOYBEAN CLONE ID:
            Gm-c1052-5141 5' similar to TR:Q9SM68 Q9SM68 ZHB0017.1. ;, mRNA
            sequence.
ACCESSION   BQ253175
VERSION     BQ253175.1  GI:20449051
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 504)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,
            Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
            Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
            Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
            Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            When it has been determined, an EST from the other end of this
            clone is listed in the 'Other ESTs on clone' field. This clone is
            available through: Biogenetic Services, 801 32nd Ave. Brookings, SD
            57006 USA (phone: 800 423 4163; email: info@biogeneticservices.com)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 420.
FEATURES              Location/Qualifiers
     source           1. .504
                      /organism="Glycine max"
                      /mol_type="mRNA"
                      /cultivar="Harosoy"
                      /db_xref="taxon:3847"
                      /clone="SOYBEAN CLONE ID: Gm-c1052-5141"
                      /tissue_type="whole seedlings of greenhouse grown plants"
                      /dev_stage="1 week old"
                      /lab_host="DH10B"
                      /clone_lib="Gm-c1052"
                      /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                      XhoI; The Harosoy NIL was constructed and seed was
                      provided by Dr. J. Specht, University of Nebraska
                      (Shoemaker and Specht, 1995). The cDNA library was
                      constructed from mRNA isolated from whole seedlings of 1
                      week old greenhouse grown plants. Complementary DNA was
                      synthesized from mRNA using a primer consisting of a
                      poly(dT) sequence with a XhoI restriction site and a 3'
                      anchor. EcoRI adapters were ligated to the blunt-ended
                      cDNA fragments followed by XhoI digestion. The cDNA
                      fragments were directionally cloned into the EcoRI-XhoI
                      restriction site of the pBluescript vector. The ligated
                      cDNA fragments were transformed into DH10B host cells
                      (GibcoBRL). The library was constructed in cooperation
                      with Dr. Paul Keim's laboratory at Northern Arizona
                      University."
ORIGIN

Alignment Scores:
Pred. No.:                2.92e+03        Length:         504
Score:                    42.00           Matches:        6
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       3               Gaps:           0

US-09-858-332G-28 (1-6) x BQ253175 (1-504)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        478 CACAACCACAACCACAAT 495

RESULT 564
CK618125
LOCUS       CK618125                 504 bp    mRNA    linear   EST 26-JAN-2004
DEFINITION  ml05h06.y1 Mouse retina, unamplified: mk/ml Mus musculus cDNA clone
            ml05h06 5', mRNA sequence.
ACCESSION   CK618125
VERSION     CK618125.1  GI:41339011
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 504)
  AUTHORS   Ida,H., Boylan,S., Weigel,A., Smit-McBride,Z., Chao,A., Gao,J.,
            Buchoff,P., Wistow,G. and Hjelmeland,L.
  TITLE     Expressed sequence tag analysis of mouse retina
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Wistow G
            Section on Molecular Structure and Function
            National Eye Institute
            6/331, NIH, Bethesda, MD 20892-2740, USA
            Tel: 301 402 3452
            Fax: 301 496 0078
            Email: graeme@helix.nih.gov
            Plate: 05  row: h  column: 06
            Seq primer: M13RP1 reverse primer (ABI).
FEATURES              Location/Qualifiers
     source           1. .504
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /strain="C57Bl6J"
                      /db_xref="taxon:10090"
                      /clone="ml05h06"
                      /sex="Male"
                      /tissue_type="retina"
```

/dev_stage="Adult"
/lab_host="EMDH10B"
/clone_lib="Mouse retina, unamplified: mk/ml"
/note="Organ: Eye; Vector: pSport1; 270ug total RNA was extracted from 200 adult male mouse retinas. A directionally cloned cDNA library in the pSPORT1 vector(Life Technologies) was constructed at Bioserve Biotechnology (Laurel MD) essentially following the protocols of the SuperScript Plasmid System full details of which are contained in the manufacturer's Instruction manual (http://www.lifetech.com/). First strand synthesis was carried out using a Not I primer-adapter [5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC(T)15-3']. cDNA was cloned in Not I/Sal I sites. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

ORIGIN

Alignment Scores:

Pred. No.:	2.92e+03	Length:	504
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	7	Gaps:	0

US-09-858-332G-28 (1-6) x CK618125 (1-504)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        109 CACAACCACAACCACAAC 126
```

RESULT 565
BM406192/c
LOCUS BM406192 505 bp mRNA linear EST 10-MAR-2003
DEFINITION EST580519 potato roots Solanum tuberosum cDNA clone cPRO27A19 5' end, mRNA sequence.
ACCESSION BM406192
VERSION BM406192.1 GI:18257822
KEYWORDS EST.
SOURCE Solanum tuberosum (potato)
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; lamiids; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 505)
AUTHORS van der Hoeven,R., Sun,H., Karamycheva,S.A., Tsai,J., Van Aken,S., Utterback,T., Chiemingo,A., Bougri,O., Buell,C.R., Ronning,C., Tanksley,S. and Baker,B.
TITLE Generation of ESTs from potato roots
JOURNAL Unpublished (2001)
COMMENT Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato-array@tigr.org
This clone can be obtained from the University of Arizona Genomics Institute. Orders can be made through URL:
http://genome.arizona.edu/orders/
Seq primer: T3.
FEATURES Location/Qualifiers
source 1. .505
/organism="Solanum tuberosum"
/mol_type="mRNA"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="cPRO27A19"
/tissue_type="roots"
/dev_stage="in vitro grown stem cuttings"
/lab_host="SOLR"
/clone_lib="potato roots"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; supplier: Cornell University, Tanksley lab; sequencing; The Institute for Genomic Research. Roots were isolated from in vitro grown stem cuttings on CM medium. Roots were isolated two weeks after placing the stem cuttings from in vitro grown plants on medium."

ORIGIN

Alignment Scores:

Pred. No.:	2.92e+03	Length:	505
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	3	Gaps:	0

US-09-858-332G-28 (1-6) x BM406192 (1-505)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        442 CATAATCATAATCATAAT 425
```

RESULT 566
BY248500
LOCUS BY248500 505 bp mRNA linear EST 10-DEC-2002
DEFINITION BY248500 RIKEN full-length enriched, visual cortex Mus musculus cDNA clone K230317L16 5', mRNA sequence.
ACCESSION BY248500
VERSION BY248500.1 GI:26430012
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 505)
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
PUBMED 12466851
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,

Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Tissues were provided by Michela Fagiolini and Takao K. Hensch (Laboratory for Neuronal Circuit Development Brain Science Institute RIKEN 2-1 Hirosawa,Wako-shi,Saitama 351-0198 Japan) whose assistance we gratefully acknowledge. Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

FEATURES Location/Qualifiers
source 1. .505
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="K230317L16"
/tissue_type="visual cortex"
/clone_lib="RIKEN full-length enriched, visual cortex"

ORIGIN

Alignment Scores:

Pred. No.:	2.92e+03	Length:	505
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	5	Gaps:	0

US-09-858-332G-28 (1-6) x BY248500 (1-505)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        305 CACAACCATAACCACAAC 322
```

RESULT 567
CA542334
LOCUS CA542334 505 bp mRNA linear EST 19-NOV-2002
DEFINITION C0617F12-5N NIA Mouse Trophoblast Stem Cell cDNA Library (Long) Mus musculus cDNA clone NIA:C0617F12 IMAGE:30021863 5', mRNA sequence.
ACCESSION CA542334
VERSION CA542334.1 GI:25084907
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 505)
AUTHORS Piao,Y., Kargul,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Luo,A., Tanaka,T., Kunath,T., Rossant,J. and Ko,M.S.H.
TITLE Systematic Analyses of NIA Mouse Trophoblast Stem Cell cDNA Library (Long)
JOURNAL Unpublished (2001)
COMMENT Other_ESTs: C0617F12-3
Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@lgsun.grc.nia.nih.gov
Plate: C0617 row: F column: 12
Seq primer: M13 Reverse
High quality sequence stop: 505
POLYA=No.

FEATURES Location/Qualifiers
source 1. .505
/organism="Mus musculus"
/mol_type="mRNA"
/strain="B5/EGFP transgenic ICR mice"
/db_xref="niaEST:C0617F12-5N"
/db_xref="taxon:10090"
/clone="NIA:C0617F12 IMAGE:30021863"
/tissue_type="Trophoblast stem cell"
/dev_stage="3.5-dpc"
/lab_host="DH10B"
/clone_lib="NIA Mouse Trophoblast Stem Cell cDNA Library (Long)"
/note="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were obtained from Dr. Janet Rossant and Tilo Kunath (Samuel Lunenfeld Research Institute, Canada). Double-stranded cDNAs were synthesized with an Oligo(dT) primer [Invitrogen: 5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTT-3'] from 4 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to Lone-linker LL-Sal4, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.6 kb. The library was constructed by Yulan Piao (NIA)."

ORIGIN

Alignment Scores:

Pred. No.:	2.92e+03	Length:	505
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-09-858-332G-28 (1-6) x CA542334 (1-505)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        257 CACAACCATAACCACAAC 274
```

RESULT 568
CF963876
LOCUS CF963876 505 bp mRNA linear EST 15-JAN-2004
DEFINITION 14637rsiced_19454.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA sequence.
ACCESSION CF963876
VERSION CF963876.1 GI:38479787
KEYWORDS EST.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

```
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 505)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
  PUBMED    15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 505
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .505
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="LYP9"
                     /db_xref="taxon:39946"
                     /tissue_type="whole plant"
                     /dev_stage="trefoil"
                     /clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA
                     library"
ORIGIN

Alignment Scores:
Pred. No.:              2.92e+03        Length:         505
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     7               Gaps:           0

US-09-858-332G-28 (1-6) x CF963876 (1-505)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        213 CACAACCACAACCACAAC 230

RESULT 569
AA049613
LOCUS       AA049613                 506 bp    mRNA    linear   EST 09-SEP-1996
DEFINITION  mj36b06.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:478163 5' similar to gb:M38188 OVARIAN GRANULOSA CELL
            13.0 KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION   AA049613
VERSION     AA049613.1  GI:1529284
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 506)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:288907
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 468.
FEATURES             Location/Qualifiers
     source          1. .506
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:478163"
                     /sex="unknown"
                     /tissue_type="embryo"
                     /dev_stage="13.5-14.5dpc total fetus"
                     /lab_host="DH10B"
                     /clone_lib="Soares mouse embryo NbME13.5 14.5"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTTT
                     T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                     14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                     State Univ., from 2 ]; double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector. Library went through one round of
                     normalization, and was constructed by Bento Soares and
                     M.Fatima Bonaldo.  "
ORIGIN

Alignment Scores:
Pred. No.:              2.93e+03        Length:         506
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AA049613 (1-506)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        108 CACAACCATAACCACAAC 125

RESULT 570
BJ365024
LOCUS       BJ365024                 506 bp    mRNA    linear   EST 08-MAR-2002
DEFINITION  BJ365024 Dictyostelium discoideum cDNA library, CF Dictyostelium
            discoideum cDNA clone ddc33i24 5', mRNA sequence.
ACCESSION   BJ365024
VERSION     BJ365024.1  GI:19274324
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 506)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the culmination
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
```

```
     source          1. .506
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddc33i24"
                     /sex="mat A"
                     /dev_stage="Culmination stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, CF"
ORIGIN

Alignment Scores:
Pred. No.:              2.93e+03       Length:         506
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     3              Gaps:           0

US-09-858-332G-28 (1-6) x BJ365024 (1-506)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        472 CACAACCACAACCACAAC 489

RESULT 571
AV541062/c
LOCUS       AV541062                507 bp    mRNA    linear   EST 20-FEB-2004
DEFINITION  AV541062 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
            cDNA clone RZ159e05F 3', mRNA sequence.
ACCESSION   AV541062
VERSION     AV541062.1  GI:8702820
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 507)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     A large scale analysis of cDNA in Arabidopsis thaliana: Generation
            of 12,028 non-redundant expressed sequence tags from normalized and
            size-selected cDNA libraries
  JOURNAL   DNA Res. 7 (3), 175-180 (2000)
   PUBMED   10907847
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .507
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /ecotype="Columbia"
                     /db_xref="taxon:3702"
                     /clone="RZ159e05F"
                     /tissue_type="roots"
                     /clone_lib="Arabidopsis thaliana roots Columbia"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
ORIGIN

Alignment Scores:
Pred. No.:              2.94e+03       Length:         507
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     1              Gaps:           0

US-09-858-332G-28 (1-6) x AV541062 (1-507)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        294 CACAACCACAACCACAAC 277

RESULT 572
BE785726/c
LOCUS       BE785726                507 bp    mRNA    linear   EST 20-OCT-2000
DEFINITION  601475292F1 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:3878087 5',
            mRNA sequence.
ACCESSION   BE785726
VERSION     BE785726.1  GI:10206924
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 507)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP/Gazdar
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9641  row: h  column: 24
            High quality sequence stop: 441.
FEATURES             Location/Qualifiers
     source          1. .507
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3878087"
                     /tissue_type="large cell carcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_68"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.8 kb.   Library constructed by Life
                     Technologies. "
ORIGIN

Alignment Scores:
Pred. No.:              2.94e+03       Length:         507
Score:                  42.00          Matches:        6
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     2              Gaps:           0

US-09-858-332G-28 (1-6) x BE785726 (1-507)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        369 CACAACCACAACCACAAC 352

RESULT 573
BJ565074/c
LOCUS       BJ565074                507 bp    mRNA    linear   EST 18-DEC-2002
DEFINITION  BJ565074 Ipomoea nil mixture of flower and flower bud Ipomoea nil
            cDNA clone jm7c07 5', mRNA sequence.
ACCESSION   BJ565074
VERSION     BJ565074.1  GI:27246894
KEYWORDS    EST.
SOURCE      Ipomoea nil (Japanese morning glory)
  ORGANISM  Ipomoea nil
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
```

```
            asterids; lamiids; Solanales; Convolvulaceae; Ipomoeeae; Ipomoea.
REFERENCE   1  (bases 1 to 507)
  AUTHORS   Hoshino,A., Seki,M., Shin-i,T., Carninci,P., Kamiya,A., Shiraki,T.,
            Nitasaka,E., Shinozaki,K., Hayashizaki,Y., Kohara,Y. and Iida,S.
  TITLE     ESTs of Japanese morning glory
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .507
                     /organism="Ipomoea nil"
                     /mol_type="mRNA"
                     /cultivar="Tokyo-kokei standard"
                     /db_xref="taxon:35883"
                     /clone="jm7c07"
                     /tissue_type="mixture of flower and flower bud"
                     /clone_lib="Ipomoea nil mixture of flower and flower bud"
ORIGIN

Alignment Scores:
Pred. No.:              2.94e+03        Length:         507
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-858-332G-28 (1-6) x BJ565074 (1-507)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        125 CACAACCACAACCACAAC 108

RESULT 574
BQ472467/c
LOCUS       BQ472467                 507 bp    mRNA    linear   EST 31-MAY-2002
DEFINITION  HB09H21T BC Hordeum vulgare subsp. vulgare cDNA clone HB09H21
            5-PRIME, mRNA sequence.
ACCESSION   BQ472467
VERSION     BQ472467.1  GI:21284596
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 507)
  AUTHORS   Radchuk,V., Zhang,H., Weschke,W., Potokina,E. and Wobus,U.
  TITLE     Barley ESTs from developing seeds
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Stein Nils
            Molecular Markers Group, Department Genbank
            Institute of Plant Genetics and Crop Plant Research (IPK)
            Corrensstr. 3, 06466, Gatersleben, Germany
            Tel: 039482-5522
            Fax: 039482-5595
            Email: stein@ipk-gatersleben.de
            Insert Length: 507   Std Error: 0.00
            Plate: 9  row: H  column: 21
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .507
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Barke"
                     /sub_species="vulgare"
                     /db_xref="taxon:112509"
                     /clone="HB09H21"
                     /tissue_type="developing caryopsis"
                     /dev_stage="8-15 DAP (days after pollination)"
                     /lab_host="XL10-Gold"
                     /clone_lib="BC"
                     /note="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of
                     cDNA); Site_2: XhoI (3'-end of cDNA); developing
                     caryopsis, 8-15 DAP(days after pollination) Due to a
                     cloning artefact caused by the kit, in most cases the
                     EcoRI site is NOT present, as well as the EcoRIadapter
                     used for cloning. To excise the insert, restriction sites
                     upstream EcoRI should be used (e.g. BamHI, SalI,PstI).
                     NOTE: Also due to the cloning system used Blue/white
                     selection for recombinats is not 100% reliable."
ORIGIN

Alignment Scores:
Pred. No.:              2.94e+03        Length:         507
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     5               Gaps:           0

US-09-858-332G-28 (1-6) x BQ472467 (1-507)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        400 CATAACCATAATCATAAC 383

RESULT 575
AI490297
LOCUS       AI490297                 508 bp    mRNA    linear   EST 18-MAY-2001
DEFINITION  EST248623 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
            cLED24J20, mRNA sequence.
ACCESSION   AI490297
VERSION     AI490297.1  GI:4385607
KEYWORDS    EST.
SOURCE      Lycopersicon esculentum (Solanum lycopersicum)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 508)
  AUTHORS   Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T.,
            Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B.,
            Fujii,C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C.,
            Martin,G.B., Tanksley,S.D. and Giovannoni,J.
  TITLE     Generation of ESTs from tomato carpel tissue
  JOURNAL   Unpublished (1999)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html.
FEATURES             Location/Qualifiers
     source          1. .508
                     /organism="Lycopersicon esculentum"
                     /mol_type="mRNA"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cLED24J20"
                     /tissue_type="carpel"
                     /dev_stage="5 days pre-anthesis to 5 days post-anthesis"
                     /lab_host="XL1-Blue MRF'"
                     /clone_lib="tomato ovary, TAMU"
                     /note="Vector: pBlueScript SK(-); Site_1: EcoR1; Site_2:
                     Xho1; cLED - Tomato Carpel EST Library. OligodT-primed and
                     directionally cloned  cDNA in vector Lamda ZAP II with 5'
                     and 3' ends located at the EcoRI and  XhoI sites,
                     respectively."
ORIGIN

Alignment Scores:
```

Pred. No.:	2.94e+03	Length:	508
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0

US-09-858-332G-28 (1-6) x AI490297 (1-508)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        260 CACAACCACAATCACAAC 277
```

RESULT 576
AL368573/c
LOCUS AL368573 508 bp mRNA linear EST 03-AUG-2000
DEFINITION MtBA25C11R1 MtBA Medicago truncatula cDNA clone MtBA25C11 T7, mRNA sequence.
ACCESSION AL368573
VERSION AL368573.1 GI:9668326
KEYWORDS EST.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE 1 (bases 1 to 508)
AUTHORS Journet,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jaillon,O., Niebel,A., Carreau,V., Chatagnier,O., Kahn,D., Gianinazzi-Pearson,V. and Gamas,P.
TITLE Medicago truncatula ESTs from nitrogen-starved roots
JOURNAL Unpublished (2000)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de Biologie Moleculaire des Relations Plantes-Microorganismes, CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email : Mt-est@toulouse.inra.fr Website : http://sequence.toulouse.inra.fr/Mtruncatula.html).
FEATURES Location/Qualifiers
source 1. .508
/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="Jemalong"
/db_xref="taxon:3880"
/clone="MtBA25C11"
/tissue_type="root tips"
/dev_stage="harvested after 3 days of N-starvation"
/clone_lib="MtBA"
/note="Vector: pBluescript pSK; Site_1: EcoRI; Site_2: XhoI; Plants were grown in an aeroponic chamber for 14 days on nitrogen-rich medium followed by 3 days on N-free medium. RNA was extracted from root tips (1-3 cm). cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into Uni-zapXR vector from Stratagene and packaged using Gigapack Gold packaging extracts. Plasmids containing cDNA inserts were mass-excised from phage stocks using ExAssit helper phage and propagated in SOLR cells. Clone ordering and sequencing was performed by the Centre National de Sequencage (Genoscope, Evry, France)."
ORIGIN

Pred. No.:	2.94e+03	Length:	508
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0

US-09-858-332G-28 (1-6) x AL368573 (1-508)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        450 CATAATCATAATCATAAT 433
```

RESULT 577
CF605814
LOCUS CF605814 508 bp mRNA linear EST 30-SEP-2003
DEFINITION RADIC01_001570 Grape Root pSPORT1 Library Vitis vinifera cDNA 5', mRNA sequence.
ACCESSION CF605814
VERSION CF605814.1 GI:37186461
KEYWORDS EST.
SOURCE Vitis vinifera
ORGANISM Vitis vinifera
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; Vitaceae; Vitis.
REFERENCE 1 (bases 1 to 508)
AUTHORS Moser,C., Segala,C., Fontana,P., Salakhudtinov,I., Gatto,P., Pindo,M., Zyprian,E., Toepfer,R., Grando,M.S. and Velasco,R.
TITLE Expressed sequence tags from different organs of Vitis vinifera
JOURNAL Unpublished (2003)
COMMENT Contact: Moser C
Laboratorio di Genetica Molecolare
Istituto Agrario di San Michele all'Adige (IASMA)
via E. Mach 1, San Michele all'Adige (TN), I-38010, Italia
Tel: 0039-0461-615314
Fax: 0039-0461-650956
Email: claudio.moser@ismaa.it
The sequencing work has been funded by the 'Fondazione Cassa di Risparmio di Trento e Rovereto'
High quality sequence stop: 508.
FEATURES Location/Qualifiers
source 1. .508
/organism="Vitis vinifera"
/mol_type="mRNA"
/cultivar="Pinot Noir"
/db_xref="taxon:29760"
/sex="Hermaphrodite"
/dev_stage="young root (first year)"
/lab_host="DH10B"
/clone_lib="Grape Root pSPORT1 Library"
/note="Organ: root; Vector: pSPORT1; Site_1: NotI; Site_2: SalI"
ORIGIN

Pred. No.:	2.94e+03	Length:	508
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-09-858-332G-28 (1-6) x CF605814 (1-508)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        128 CACAACCACAACCACAAC 145
```

RESULT 578
CF991731
LOCUS CF991731 508 bp mRNA linear EST 04-FEB-2005
DEFINITION 17944rsicee_8818.y1 Oryza sativa cv. LYP9 tillering whole plant cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA sequence.
ACCESSION CF991731
VERSION CF991731.1 GI:58593423
KEYWORDS EST.
SOURCE Oryza sativa (indica cultivar-group)

```
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 508)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 508
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .508
                      /organism="Oryza sativa (indica cultivar-group)"
                      /mol_type="mRNA"
                      /cultivar="LYP9"
                      /db_xref="taxon:39946"
                      /tissue_type="whole plant"
                      /dev_stage="tillering"
                      /clone_lib="Oryza sativa cv. LYP9 tillering whole plant
                      cDNA library"
ORIGIN

Alignment Scores:
Pred. No.:              2.94e+03        Length:         508
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     7               Gaps:           0

US-09-858-332G-28 (1-6) x CF991731 (1-508)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        219 CACAACCACAACCACAAC 236

RESULT 579
AA636864
LOCUS       AA636864                509 bp    mRNA    linear   EST 22-OCT-1997
DEFINITION  vr20g04.r1 Barstead mouse myotubes MPLRB5 Mus musculus cDNA clone
            IMAGE:1121238 5' similar to gb:M38188 OVARIAN GRANULOSA CELL 13.0
            KD PROTEIN HGR74 (HUMAN);, mRNA sequence.
ACCESSION   AA636864
VERSION     AA636864.1  GI:2560643
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 509)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:610574
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 323.
FEATURES              Location/Qualifiers
     source           1. .509
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /strain="C3H"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:1121238"
                      /cell_line="C2C12"
                      /lab_host="DH10B"
                      /clone_lib="Barstead mouse myotubes MPLRB5"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
                      was primed with a Not I - oligo(dT) primer [5'
                      TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTT
                      3']; double-stranded cDNA was ligated to Eco RI adaptors
                      [AATTCGGATCCTTG], digested with Not I and cloned into the
                      Not I and Eco RI sites of the modified pT7T3 vector.
                      Library constructed by Bob Barstead. The C2C12 cell line
                      (available from ATCC, catalog # CRL-1772) differentiates
                      rapidly, forming contractile myotubes and producing
                      characteristic muscle proteins. "
ORIGIN

Alignment Scores:
Pred. No.:              2.95e+03        Length:         509
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AA636864 (1-509)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        256 CACAACCATAACCACAAC 273

RESULT 580
BF177637
LOCUS       BF177637                509 bp    mRNA    linear   EST 31-OCT-2000
DEFINITION  Ljirnpest32-586-g12 Ljirnp Lambda HybriZap two-hybrid library Lotus
            corniculatus var. japonicus cDNA clone LP586-32-g12 5' similar to
            mRNA for integral membrane protein, probably sugar transporter,
            mRNA sequence.
ACCESSION   BF177637
VERSION     BF177637.1  GI:11065328
KEYWORDS    EST.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1  (bases 1 to 509)
  AUTHORS   Poulsen,C. and Poedenphant,L.
  TITLE     Expressed sequence tags from Mesorhizobium loti infected roots of
            Lotus japonicus
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Poulsen,C.; Poedenphant,L.
            Laboratory of Gene Expression, Department of Molecular and
            Structural Biology
```

```
            University of Aarhus
            Gustav Wieds Vej 10C, DK-8000 Aarhus C, Denmark
            Tel: +45 89425007
            Fax: +45 86201222
            Email: CHP@mbio.aau.dk
            Seq primer: pADGAl4 5'-primer CCA CTA CAA TGG ATG ATG TAT ATA AC
            High quality sequence stop: 546.
FEATURES              Location/Qualifiers
     source           1. .509
                      /organism="Lotus corniculatus var. japonicus"
                      /mol_type="mRNA"
                      /cultivar="variety Gifu B-129"
                      /db_xref="taxon:34305"
                      /clone="LP586-32-g12"
                      /dev_stage="young plants"
                      /lab_host="Escherichia coli XL-1Blue MRF'"
                      /clone_lib="Ljirnp Lambda HybriZap two-hybrid library"
                      /note="Organ: whole roots and root nodule primordia;
                      Vector: pADGAL4; Site_1: EcoRI; Site_2: XhoI; Two times
                      polyA+ purified mRNA, oligodT primed, 5' adapter tag
                      GAATTCGGCACGAG "
ORIGIN

Alignment Scores:
Pred. No.:                2.95e+03        Length:          509
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       2               Gaps:            0

US-09-858-332G-28 (1-6) x BF177637 (1-509)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         67 CATAATCATAATCATAAT 84

RESULT 581
CF991173
LOCUS       CF991173                509 bp    mRNA    linear   EST 04-FEB-2005
DEFINITION  18042rsicee_8929.y1 Oryza sativa cv. LYP9 tillering whole plant
            cDNA library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF991173
VERSION     CF991173.1  GI:58592865
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 509)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
            High quality sequence stop: 509
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .509
                      /organism="Oryza sativa (indica cultivar-group)"
                      /mol_type="mRNA"
                      /cultivar="LYP9"
                      /db_xref="taxon:39946"
                      /tissue_type="whole plant"
                      /dev_stage="tillering"
                      /clone_lib="Oryza sativa cv. LYP9 tillering whole plant
                      cDNA library"
ORIGIN

Alignment Scores:
Pred. No.:                2.95e+03        Length:          509
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       7               Gaps:            0

US-09-858-332G-28 (1-6) x CF991173 (1-509)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        207 CACAACCACAACCACAAC 224

RESULT 582
CO333305/c
LOCUS       CO333305                509 bp    mRNA    linear   EST 29-JUN-2004
DEFINITION  EK306440.5prime Exelixis FlyTag CK01 pCDNA-SK+ Drosophila
            melanogaster cDNA clone EK306440 5, mRNA sequence.
ACCESSION   CO333305
VERSION     CO333305.1  GI:49393580
KEYWORDS    EST.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 509)
  AUTHORS   Kopczynski,C., Platt,D., Campbell,J., Muzong,C., Laufer,A.,
            Peterson,E. and Swimmer,C.
  TITLE     Exelixis FlyTag EST Project CK01 Library
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: EK.3064  row: D  column: 4
            High quality sequence stop: 455.
FEATURES              Location/Qualifiers
     source           1. .509
                      /organism="Drosophila melanogaster"
                      /mol_type="mRNA"
                      /db_xref="taxon:7227"
                      /clone="EK306440"
                      /clone_lib="Exelixis FlyTag CK01 pCDNA-SK+"
                      /note="Organ: mixed stage embryos, imaginal disks, and
                      adult heads; Vector: pCDNA-SK+; Site_1: NotI; Site_2:
                      XhoI; Random primed, normalized library from mixed stage
                      embryos, imaginal disks, and adult heads."
ORIGIN

Alignment Scores:
Pred. No.:                2.95e+03        Length:          509
Score:                    42.00           Matches:         6
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       7               Gaps:            0

US-09-858-332G-28 (1-6) x CO333305 (1-509)
```

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        480 CATAATCATAATCATAAT 463

RESULT 583
CX535329
LOCUS       CX535329                509 bp    mRNA    linear   EST 11-JAN-2005
DEFINITION  s13dNF39G03MJ024_388827 Methyl Jasmonate-Elicited Root Cell
            Suspension Culture Medicago truncatula cDNA, mRNA sequence.
ACCESSION   CX535329
VERSION     CX535329.1  GI:57504048
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 509)
  AUTHORS   Achnine,L., He,X.-Z., Scott,A.D., May,G.D. and Dixon,R.A.
  TITLE     Medicago truncatula Methyl Jasmonate-Elicited Root Cell Suspension
            Culture Expressed Sequence Tags from the Samuel Roberts Noble
            Foundation - Center for Medicago Genomics Research
  JOURNAL   Unpublished (2005)
COMMENT     Contact: May GD
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA
            Tel: 580 224 6650
            Fax: 580 224 6692
            Email: gdmay@noble.org.
FEATURES             Location/Qualifiers
     source          1. .509
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /db_xref="taxon:3880"
                     /tissue_type="Root cell suspension culture"
                     /dev_stage="Seven-day old culture"
                     /clone_lib="Methyl Jasmonate-Elicited Root Cell Suspension
                     Culture"
                     /note="Vector: Lambda Zap; Medicago truncatula root cell
                     suspension culture elicited with 500 um methyl jasmonate.
                     The cDNA was directionally ligated into the Uni-Zap XR
                     vector (Stratagene) and packaged using the Gigapack III
                     Gold packaging extracts. Phagemids containing cDNA inserts
                     were in vivo excised from the recombinant Uni-ZAP XR
                     vector using ExAssist helper phage and the E. coli strain
                     XL1-Blue MRF' (Stratagene). Excised plasmids were plated
                     using SOLR cells."
ORIGIN

Alignment Scores:
Pred. No.:              2.95e+03        Length:         509
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     8               Gaps:           0

US-09-858-332G-28 (1-6) x CX535329 (1-509)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         84 CACAATCACAATCACAAT 101

RESULT 584
CA890073/c
LOCUS       CA890073                510 bp    mRNA    linear   EST 20-DEC-2002
DEFINITION  B0156H06-5N NIA Mouse Neural Stem Cell (Differentiated) cDNA
            Library (Long) Mus musculus cDNA clone NIA:B0156H06 IMAGE:30099257
            5', mRNA sequence.
ACCESSION   CA890073
VERSION     CA890073.1  GI:27341622
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 510)
  AUTHORS   Piao,Y., Dudekula,D.B., Qian,Y., Martin,P.R., Aiba,K., Vescovi,A.L.
            and Ko,M.S.H.
  TITLE     Systematic Analyses of NIA Mouse Neural Stem Cell (Differentiated)
            cDNA Library (Long)
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Dawood B. Dudekula
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgsun.grc.nia.nih.gov
            Plate: B0156  row: H  column: 06
            Seq primer: -21M13 Reverse
            High quality sequence stop: 510
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .510
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="CD1"
                     /db_xref="niaEST:B0156H06-5N"
                     /db_xref="taxon:10090"
                     /clone="NIA:B0156H06 IMAGE:30099257"
                     /dev_stage="Adult"
                     /lab_host="DH10B"
                     /clone_lib="NIA Mouse Neural Stem Cell (Differentiated)
                     cDNA Library (Long)"
                     /note="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2:
                     NotI; Mouse cDNA project by the Laboratory of Genetics,
                     National Institute on Aging (NIA), Intramural Research
                     Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is
                     a long-transcript enriched cDNA library (Ref. Genome Res.
                     11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were
                     obtained from Dr. Angelo L. Vescovi (Institute for Stem
                     Cell Research, Italy). Double-stranded cDNAs were
                     synthesized with an Oligo(dT) primer [Invitrogen:
                     5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTTT-3'] from
                     2.0 Microgram of total RNA, treated with T4 DNA
                     polymerase, and purified by ethanol-precipitation. The
                     cDNAs were ligated to Lone-linker LL-Sal4, purified by
                     phenol/chloroform, and separated from free linkers by
                     Centricon 100. Then, the cDNAs were amplified by
                     long-range high fidelity PCR using Ex Taq polymerase
                     (Takara) with a primer Sal4-S. The products were purified
                     by phenol/chloroform and Centricon 100. The cDNAs were
                     digested with SalI and NotI enzymes and cloned into
                     SalI/NotI site of pSPORT1 plasmid vector. The DH10B E.
                     coli host was transformed with the ligation mixture by the
                     standard chemical method. The average insert size is about
                     3.2 kb. The library was constructed by Yulan Piao."
ORIGIN

Alignment Scores:
Pred. No.:              2.96e+03        Length:         510
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x CA890073 (1-510)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        151 CACAATCACAACCACAAT 134
```

```
RESULT 585
BG060120
LOCUS       BG060120                511 bp    mRNA    linear   EST 25-JAN-2001
DEFINITION  naf49b10.x1 NCI_CGAP_Brn65 Homo sapiens cDNA clone IMAGE:4147171
            3', mRNA sequence.
ACCESSION   BG060120
VERSION     BG060120.1  GI:12528281
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 511)
  AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David N. Louis, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer: -40UP from Gibco
            High quality sequence stop: 436.
FEATURES             Location/Qualifiers
     source          1. .511
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4147171"
                     /tissue_type="glioblastoma without EGFR amplification"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Brn65"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.77 kb. Constructed by Life
                     Technologies."
ORIGIN

Alignment Scores:
Pred. No.:                2.96e+03       Length:          511
Score:                    42.00          Matches:         6
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       2              Gaps:            0

US-09-858-332G-28 (1-6) x BG060120 (1-511)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        195 CATAATCATAATCATAAT 212

RESULT 586
BJ344984/c
LOCUS       BJ344984                511 bp    mRNA    linear   EST 06-MAR-2002
DEFINITION  BJ344984 Dictyostelium discoideum cDNA library, AF Dictyostelium
            discoideum cDNA clone dda21p04 3', mRNA sequence.
ACCESSION   BJ344984
VERSION     BJ344984.1  GI:19215491
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 511)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the aggregation
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .511
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dda21p04"
                     /sex="mat A"
                     /dev_stage="Aggregation stage"
                     /clone_lib="Dictyostelium discoideum cDNA library, AF"
ORIGIN

Alignment Scores:
Pred. No.:                2.96e+03       Length:          511
Score:                    42.00          Matches:         6
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       3              Gaps:            0

US-09-858-332G-28 (1-6) x BJ344984 (1-511)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        468 CACAATCACAACCACAAC 451

RESULT 587
BQ985398
LOCUS       BQ985398                511 bp    mRNA    linear   EST 21-AUG-2002
DEFINITION  QGE6L08.yg.ab1 QG_EFGHJ lettuce serriola Lactuca sativa cDNA clone
            QGE6L08, mRNA sequence.
ACCESSION   BQ985398
VERSION     BQ985398.1  GI:22402923
KEYWORDS    EST.
SOURCE      Lactuca sativa
  ORGANISM  Lactuca sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
            Cichorieae; Lactuca.
REFERENCE   1  (bases 1 to 511)
  AUTHORS   Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J.,
            Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y.,
            Lai,Z., Church,S., Jackson,L. and Bradford,K.
  TITLE     Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
            belongs to contig QG_CA_Contig3547, see http://cgpdb.ucdavis.edu/
            for details.
            Plate: QGE6  row: L  column: 08.
FEATURES             Location/Qualifiers
     source          1. .511
                     /organism="Lactuca sativa"
                     /mol_type="mRNA"
                     /cultivar="L.serriola"
                     /db_xref="taxon:4236"
```

/clone="QGE6L08"
/lab_host="E.coli"
/clone_lib="QG_EFGHJ lettuce serriola"
/note="Vector: pBRcDNASfiAB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at http://cgpdb.ucdavis.edu/ TAG_SEQ=Not found"

ORIGIN

Alignment Scores:

Pred. No.:	2.96e+03	Length:	511
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	5	Gaps:	0

US-09-858-332G-28 (1-6) x BQ985398 (1-511)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        248 CACAATCACAATCACAAC 265
```

RESULT 588
CD340364
LOCUS CD340364 511 bp mRNA linear EST 17-SEP-2003
DEFINITION StrPu536.004185 Sea urchin embryo 40hr gastrula stage cDNA library MPMGp536 Strongylocentrotus purpuratus cDNA clone CALTp536M0226;MPI_536_26M2 3', mRNA sequence.
ACCESSION CD340364
VERSION CD340364.1 GI:34806890
KEYWORDS EST.
SOURCE Strongylocentrotus purpuratus
ORGANISM Strongylocentrotus purpuratus
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae; Strongylocentrotus.
REFERENCE 1 (bases 1 to 511)
AUTHORS Poustka,A.J., Groth,D., Hennig,S., Thamm,S., Cameron,A., Beck,A., Reinhardt,R., Herwig,R., Panopoulou,G. and Lehrach,H.
TITLE Generation, annotation, evolutionary analysis, and database integration of 20,000 unique sea urchin EST clusters
JOURNAL Genome Res. 13 (12), 2736-2746 (2003)
COMMENT Contact: Poustka AJ
laboraty 145, dept.Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Ihnestr.63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1235
Fax: +49 30 8413 1128
Email: poustka@molgen.mpg.de
The library was characterised by oligonucleotide fingerprinting (ONF) to reduce sequencing redundancy. According to the ONF procedure, clones that display the same hybridisation matrix with a battery of 200 8mer oligonucleotides are grouped into clusters. One clone per ONF cluster is selected for sequencing. The size of each cluster is an indicator of the frequency of a transcript in the analysed library. The cluster size as well as the coordinates of the other clones assigned to the same ONF cluster as the clone from which the above EST is generated is available at the sea urchin project web site at: http://www.molgen.mpg.de/ag_seaurchin/ . cDNA clones and filters are distributed via the Resource Center/Primary Database of the German Human Genome Project (http://www.rzpd.de)
PCR PRimers
FORWARD: 5' CCCCAGGCTTTACACTTTATGCTTCCGGCTCG 3' (M13RSP) 5'-seq
BACKWARD: 5' GCTATTACGCCAGCTGGCGAAAGGGGGATGTG 3' (M13FSP) 3'-seq
Seq primer: 5' GCTATTACGCCAGCTGGCGAAAGGGGGATGTG 3' (M13FSP)
High quality sequence stop: 511.

FEATURES Location/Qualifiers
source 1. .511
/organism="Strongylocentrotus purpuratus"
/mol_type="mRNA"
/db_xref="taxon:7668"
/clone="CALTp536M0226;MPI_536_26M2"
/tissue_type="whole embryo"
/dev_stage="embryonic 40hr"
/lab_host="E.coli, XL1 blue"
/clone_lib="Sea urchin embryo 40hr gastrula stage cDNA library MPMGp536"
/note="Vector: pSport1; Site_1: NotI; Site_2: SalI; Random primed and directionally cloned in pSport1 vector using a NotI (5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC (T)15-3' and a SalI 5'- TCGACCCACGCGTCCG-3'adapters (Gibco BRL)"

ORIGIN

Alignment Scores:

Pred. No.:	2.96e+03	Length:	511
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-09-858-332G-28 (1-6) x CD340364 (1-511)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        240 CATAACCATAACCATAAC 257
```

RESULT 589
CL283244
LOCUS CL283244 511 bp DNA linear GSS 10-FEB-2004
DEFINITION ZMMBBb0629A13r ZMMBBb (HindIII) Zea mays genomic clone ZMMBBb0629A13 3', genomic survey sequence.
ACCESSION CL283244
VERSION CL283244.1 GI:42497631
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 511)
AUTHORS Bharti,A.K., Young,S., Kavchok,S., Keizer,G., Bronzino,A.C., Zohovetz,V., Fuks,G., Yu,Y., Wing,R. and Messing,J.
TITLE Sequencing of the maize genome at PGIR (2003c)
JOURNAL Unpublished (2003)
COMMENT Contact: Bharti,A.K.
Dr.Joachim Messing's lab
The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers University
190 Frelinghuysen Road, Piscataway, NJ 08854, USA
Tel: 732 445 3801
Fax: 732 445 5735
Email: bharti@waksman.rutgers.edu
Seq primer: SP6
Class: BAC ends
High quality sequence start: 85.
FEATURES Location/Qualifiers
source 1. .511
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="B73"
/db_xref="taxon:4577"
/clone="ZMMBBb0629A13"
/lab_host="E. coli DH10B"
/clone_lib="ZMMBBb (HindIII)"
/note="Vector: pCUGI; Site_1: HindIII; Site_2: HindIII"

ORIGIN

Alignment Scores:

```
Pred. No.:              2.96e+03        Length:         511
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     10              Gaps:           0

US-09-858-332G-28 (1-6) x CL283244 (1-511)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        332 CATAATCATAATCATAAT 349

RESULT 590
CR297245
LOCUS       CR297245                511 bp    DNA     linear   GSS 19-NOV-2004
DEFINITION  mte1-14A4RM1 BAC end, cultivar Jemalong A17 of Medicago truncatula,
            genomic survey sequence.
ACCESSION   CR297245
VERSION     CR297245.1  GI:44692554
KEYWORDS    GSS.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 511)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
FEATURES             Location/Qualifiers
     source          1. .511
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /cultivar="Jemalong A17"
                     /db_xref="taxon:3880"
                     /clone_lib="MTE1"
                     /note="Vector: pIndigoBAC ; Site_1: EcoRI ; Site_2: EcoRI
                     ; Debelle F. and Chalhoub B.
                     Genoscope sequence ID : mte1-14A4RM1"
ORIGIN

Alignment Scores:
Pred. No.:              2.96e+03        Length:         511
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     11              Gaps:           0

US-09-858-332G-28 (1-6) x CR297245 (1-511)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        143 CATAATCATAATCATAAT 160

RESULT 591
AW720113
LOCUS       AW720113                512 bp    mRNA    linear   EST 19-APR-2000
DEFINITION  LjNEST14b12r Lotus japonicus nodule library 5 and 7 week-old Lotus
            corniculatus var. japonicus cDNA 5', mRNA sequence.
ACCESSION   AW720113
VERSION     AW720113.1  GI:7614644
KEYWORDS    EST.
SOURCE      Lotus corniculatus var. japonicus (Lotus japonicus)
  ORGANISM  Lotus corniculatus var. japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1  (bases 1 to 512)
  AUTHORS   Colebatch,G., Freund,S., Trevaskis,B and Udvardi,M.
  TITLE     Lotus japonicus root nodule ESTs: tools for functional genomics
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Udvardi MK
            Molecular Plant Nutrition
            Max Planck Institute of Molecular Plant Physiology
            Am Muehlenberg 1, 14476 Golm, Germany
            Fax: 49 331 567 8250
            Email: udvardi@mpimp-golm.mpg.de
            Seq primer: T7
            High quality sequence stop: 512.
FEATURES             Location/Qualifiers
     source          1. .512
                     /organism="Lotus corniculatus var. japonicus"
                     /mol_type="mRNA"
                     /cultivar="Gifu (B-129)"
                     /db_xref="taxon:34305"
                     /dev_stage="5 and 7 week-old plants"
                     /clone_lib="Lotus japonicus nodule library 5 and 7
                     week-old"
                     /note="Organ: Nodule; Vector: pSPORT1; Site_1: SalI;
                     Site_2: NotI; The library was prepared using mRNA
                     extracted from nodules of 5 and 7 week-old Lotus plants.
                     Nodules were induced by, and contained Mesorhizobium
                     strain R7A."
ORIGIN

Alignment Scores:
Pred. No.:              2.97e+03        Length:         512
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AW720113 (1-512)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        137 CACAACCACAACCACAAC 154

RESULT 592
BJ384097/c
LOCUS       BJ384097                512 bp    mRNA    linear   EST 08-MAR-2002
DEFINITION  BJ384097 Dictyostelium discoideum cDNA library, CF Dictyostelium
            discoideum cDNA clone ddc48b11 3', mRNA sequence.
ACCESSION   BJ384097
VERSION     BJ384097.1  GI:19293481
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 512)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the culmination
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .512
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddc48b11"
                     /sex="mat A"
```

```
                /dev_stage="Culmination stage"
                /clone_lib="Dictyostelium discoideum cDNA library, CF"
ORIGIN

Alignment Scores:
Pred. No.:              2.97e+03        Length:         512
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-858-332G-28 (1-6) x BJ384097 (1-512)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        478 CACAATCACAACCACAAC 461

RESULT 593
CF083291
LOCUS       CF083291                512 bp    mRNA    linear   EST 22-JUL-2003
DEFINITION  QHL1A12.yg.ab1 QH_L sunflower H.paradoxus (salt stress) Helianthus
            paradoxus cDNA clone QHL1A12, mRNA sequence.
ACCESSION   CF083291
VERSION     CF083291.1  GI:33122334
KEYWORDS    EST.
SOURCE      Helianthus paradoxus
  ORGANISM  Helianthus paradoxus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; campanulids; Asterales; Asteraceae; Asteroideae;
            Heliantheae; Helianthus.
REFERENCE   1  (bases 1 to 512)
  AUTHORS   Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J.,
            Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y.,
            Lai,Z., Church,S., Jackson,L. and Bradford,K.
  TITLE     Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
            singleton, see http://cgpdb.ucdavis.edu/ for details.
            Plate: QHL1  row: A  column: 12.
FEATURES             Location/Qualifiers
     source          1. .512
                     /organism="Helianthus paradoxus"
                     /mol_type="mRNA"
                     /db_xref="taxon:73304"
                     /clone="QHL1A12"
                     /lab_host="E.coli"
                     /clone_lib="QH_L sunflower H.paradoxus (salt stress)"
                     /note="Vector: pGEM-T; The library was constructed from
                     four different sources (seedling, root, leaf and flower)
                     of RNA from a single genotype. cDNAs were pooled and
                     cloned into a high-copy vector pGEM-T. Details of library
                     construction can be obtained at http://cgpdb.ucdavis.edu/"
ORIGIN

Alignment Scores:
Pred. No.:              2.97e+03        Length:         512
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-858-332G-28 (1-6) x CF083291 (1-512)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         16 CATAATCATAACCATAAT 33

RESULT 594
CO188836
LOCUS       CO188836                512 bp    mRNA    linear   EST 21-JUN-2004
DEFINITION  EK042146.5prime Exelixis FlyTag CK01 pCDNA-SK+ Drosophila
            melanogaster cDNA clone EK042146 5, mRNA sequence.
ACCESSION   CO188836
VERSION     CO188836.1  GI:49000011
KEYWORDS    EST.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 512)
  AUTHORS   Kopczynski,C., Platt,D., Campbell,J., Muzong,C., Laufer,A.,
            Peterson,E. and Swimmer,C.
  TITLE     Exelixis FlyTag EST Project CK01 Library
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Based upon one or more reads of this clone where vector sequence
            was present at both ends, this clone has been determined to contain
            contain a cDNA insert on the order of 600-1000 bases.
            Plate: EK.421  row: D  column: 10
            High quality sequence stop: 456.
FEATURES             Location/Qualifiers
     source          1. .512
                     /organism="Drosophila melanogaster"
                     /mol_type="mRNA"
                     /db_xref="taxon:7227"
                     /clone="EK042146"
                     /clone_lib="Exelixis FlyTag CK01 pCDNA-SK+"
                     /note="Organ: mixed stage embryos, imaginal disks, and
                     adult heads; Vector: pCDNA-SK+; Site_1: NotI; Site_2:
                     XhoI; Random primed, normalized library from mixed stage
                     embryos, imaginal disks, and adult heads."
ORIGIN

Alignment Scores:
Pred. No.:              2.97e+03        Length:         512
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     7               Gaps:           0

US-09-858-332G-28 (1-6) x CO188836 (1-512)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        332 CACAATCACAACCACAAT 349

RESULT 595
BI094976/c
LOCUS       BI094976                513 bp    mRNA    linear   EST 21-JUL-2004
DEFINITION  sae02c07.y1 Gm-c1055 Glycine max cDNA clone GENOME SYSTEMS CLONE
            ID: Gm-c1055-3950 5' similar to TR:Q9SY30 Q9SY30 T17H7.16. ;, mRNA
            sequence.
ACCESSION   BI094976
VERSION     BI094976.1  GI:14516304
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
```

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE 1 (bases 1 to 513)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
When it has been determined, an EST from the other end of this clone is listed in the 'Other ESTs on clone' field. This clone is available through: Biogenetic Services, 801 32nd Ave. Brookings, SD 57006 USA (phone: 800 423 4163; email: info@biogeneticservices.com)
Seq primer: -40RP from Gibco
High quality sequence stop: 414.
FEATURES Location/Qualifiers
source 1. .513
/organism="Glycine max"
/mol_type="mRNA"
/cultivar="L82-2024"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1055-3950"
/tissue_type="Mature seed pods, greenhouse grown"
/lab_host="DH10B"
/clone_lib="Gm-c1055"
/note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2: XhoI; The cDNA library was constructed from mRNA isolated from mature seed pods of greenhouse grown plants prior to senescence for the cultivar KP1. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (GibcoBRL). This library was constructed in the laboratory of Dr. Randy Shoemaker."
ORIGIN

Alignment Scores:

Pred. No.:	2.97e+03	Length:	513
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-858-332G-28 (1-6) x BI094976 (1-513)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db         50 CACAATCACAATCACAAT 33
```

RESULT 596
CN827064/c
LOCUS CN827064 513 bp mRNA linear EST 14-MAY-2005
DEFINITION EL2007R Brassica embryo library (EL) Brassica napus cDNA clone EL2007 5', mRNA sequence.
ACCESSION CN827064
VERSION CN827064.1 GI:65296850
KEYWORDS EST.
SOURCE Brassica napus (rape)
ORGANISM Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 513)
AUTHORS Sharpe,A.G., Gjetvaj,B., Durkin,J. and Lydiate,D.J.
TITLE Brassica napus ESTs
JOURNAL Unpublished (2004)
COMMENT Contact: Sharpe, A.G.
Molecular Genetics
Agriculture & Agri-Food Canada
107 Science Place, Saskatoon, Sakatchewan, Canada, S7N0X2
Tel: 306 956 7271
Fax: 306 956 7247
Email: sharpea@agr.gc.ca
Seq primer: T7.
FEATURES Location/Qualifiers
source 1. .513
/organism="Brassica napus"
/mol_type="mRNA"
/cultivar="DH12075 (double haploid line from Cresor x Westar cross)"
/db_xref="taxon:3708"
/clone="EL2007"
/dev_stage="Mid to late embryos (4-6 mg)"
/lab_host="E. coli ElectroMAX DH5 alpha-e (Invitrogen)"
/clone_lib="Brassica embryo library (EL)"
/note="Organ: Embryos without seed coat; Vector: pSPORT1 (modified: GCGGCCGCCc*GACTAGTGAGCTC*cggacgcgtggGTCGAC); Site_1: NotI; Site_2: SalI; Seeds were collected by Dr. Francois Ouellet when they were still very green (mid to large stage, cotyledons were formed). The seed coats were removed and the remaining tissue was used for cDNA library construction. mRNA was poly-A primed using SuperScript Plasmid System cDNA Synthesis and Cloning kit (Invitrogen) After initial screening, the most abundant redundant clones were screened out using 22 oligos designed to match napins (including albumins), cruciferins, oleosins, trypsin inhibitor 2, cytosolic GAPDH, cyclophilins, HSP70, desaturase, and CAB (LHCP)."
ORIGIN

Alignment Scores:

Pred. No.:	2.97e+03	Length:	513
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	7	Gaps:	0

US-09-858-332G-28 (1-6) x CN827064 (1-513)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        135 CATAACCATAACCATAAC 118
```

RESULT 597
AV567232/c
LOCUS AV567232 514 bp mRNA linear EST 23-FEB-2004
DEFINITION AV567232 Arabidopsis thaliana green siliques Columbia Arabidopsis thaliana cDNA clone SQL09d01F 3', mRNA sequence.
ACCESSION AV567232
VERSION AV567232.1 GI:8738686
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 514)
AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and

```
          size-selected cDNA libraries
  JOURNAL   DNA Res. 7 (3), 175-180 (2000)
   PUBMED   10907847
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .514
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /ecotype="Columbia"
                     /db_xref="taxon:3702"
                     /clone="SQL09d01F"
                     /tissue_type="green siliques"
                     /clone_lib="Arabidopsis thaliana green siliques Columbia"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
ORIGIN

Alignment Scores:
Pred. No.:              2.98e+03        Length:         514
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-858-332G-28 (1-6) x AV567232 (1-514)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        297 CACAACCACAACCACAAC 280

RESULT 598
BM886283
LOCUS       BM886283                514 bp    mRNA    linear   EST 06-JUL-2004
DEFINITION  sam14f07.y1 Gm-c1068 Glycine max cDNA clone SOYBEAN CLONE ID:
            Gm-c1068-4862 5' similar to TR:O65168 O65168 AUXIN-INDUCED PROTEIN
            ;, mRNA sequence.
ACCESSION   BM886283
VERSION     BM886283.1  GI:19270027
KEYWORDS    EST.
SOURCE      Glycine max (soybean)
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 514)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,
            Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
            Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
            Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
            Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            When it has been determined, an EST from the other end of this
            clone is listed in the 'Other ESTs on clone' field. This clone is
            available through: Biogenetic Services, 801 32nd Ave. Brookings, SD
            57006 USA (phone: 800 423 4163; email: info@biogeneticservices.com)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 421.
FEATURES             Location/Qualifiers
     source          1. .514
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /cultivar="Williams 82"
                     /db_xref="taxon:3847"
                     /clone="SOYBEAN CLONE ID: Gm-c1068-4862"
                     /tissue_type="Leaf, drought stressed, 1 month old plants,
                     greenhouse grown"
                     /lab_host="DH10B"
                     /clone_lib="Gm-c1068"
                     /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library was constructed from mRNA isolated
                     from drought stressed leaf tissue of the cultivar Williams
                     82. The month old greenhouse grown plants were deprived of
                     water for 3 days prior to harvesting the stressed leaf
                     tissue. Complementary DNA was synthesized from mRNA using
                     a primer consisting of a poly(dT) sequence with a XhoI
                     restriction site. EcoRI adapters were ligated to the
                     blunt-ended cDNA fragments followed by XhoI digestion. The
                     cDNA fragments were directionally cloned into the
                     EcoRI-XhoI restriction site of the pBluescript vector. The
                     ligated cDNA fragments were transformed into DH10B host
                     cells (GibcoBRL). This library was constructed in the
                     laboratory of Dr. Randy Shoemaker."
ORIGIN

Alignment Scores:
Pred. No.:              2.98e+03        Length:         514
Score:                  42.00           Matches:        6
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-858-332G-28 (1-6) x BM886283 (1-514)

Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        293 CATAACCATAACCATAAC 310

RESULT 599
CF965313
LOCUS       CF965313                514 bp    mRNA    linear   EST 15-JAN-2004
DEFINITION  11490rsiced_3177.y1 Oryza sativa cv. LYP9 trefoil whole plant cDNA
            library Oryza sativa (indica cultivar-group) cDNA 5', mRNA
            sequence.
ACCESSION   CF965313
VERSION     CF965313.1  GI:38481224
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 514)
  AUTHORS   Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W.,
            Hu,S., Zeng,C., Zhang,J., Zhang,Y., Li,R., Xu,Z., Li,S., Li,X.,
            Zheng,H., Cong,L., Lin,L., Yin,J., Geng,J., Li,G., Shi,J., Liu,J.,
            Lv,H., Li,J., Wang,J., Deng,Y., Ran,L., Shi,X., Wang,X., Wu,Q.,
            Li,C., Ren,X., Wang,J., Wang,X., Li,D., Liu,D., Zhang,X., Ji,Z.,
            Zhao,W., Sun,Y., Zhang,Z., Bao,J., Han,Y., Dong,L., Ji,J., Chen,P.,
            Wu,S. and Liu,J.
  TITLE     The Genomes of Oryza sativa: A History of Duplications
  JOURNAL   PLoS Biol. 3 (2), e38 (2005)
   PUBMED   15685292
COMMENT     Contact: Yan Zhou
            Bioinformatics Department
            Hangzhou Genomics Institute
            No.51 Zhijiang Road, Hangzhou 310008, China
            Tel: 86-571-56805886
            Fax: 86-571-56805884
            Email: zhouyan@genomics.org.cn
            Seq primer: M13 Forward
```

High quality sequence stop: 514
POLYA=No.

FEATURES Location/Qualifiers
source 1. .514
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/cultivar="LYP9"
/db_xref="taxon:39946"
/tissue_type="whole plant"
/dev_stage="trefoil"
/clone_lib="Oryza sativa cv. LYP9 trefoil whole plant cDNA library"

ORIGIN

Alignment Scores:

Pred. No.:	2.98e+03	Length:	514
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	7	Gaps:	0

US-09-858-332G-28 (1-6) x CF965313 (1-514)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        145 CACAACCACAACCACAAC 162
```

RESULT 600
CN827673/c
LOCUS CN827673 514 bp mRNA linear EST 14-MAY-2005
DEFINITION EL2449R Brassica embryo library (EL) Brassica napus cDNA clone EL2449 5', mRNA sequence.
ACCESSION CN827673
VERSION CN827673.1 GI:65297459
KEYWORDS EST.
SOURCE Brassica napus (rape)
ORGANISM Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 514)
AUTHORS Sharpe,A.G., Gjetvaj,B., Durkin,J. and Lydiate,D.J.
TITLE Brassica napus ESTs
JOURNAL Unpublished (2004)
COMMENT Contact: Sharpe, A.G.
Molecular Genetics
Agriculture & Agri-Food Canada
107 Science Place, Saskatoon, Sakatchewan, Canada, S7N0X2
Tel: 306 956 7271
Fax: 306 956 7247
Email: sharpea@agr.gc.ca
Seq primer: T7.

FEATURES Location/Qualifiers
source 1. .514
/organism="Brassica napus"
/mol_type="mRNA"
/cultivar="DH12075 (double haploid line from Cresor x Westar cross)"
/db_xref="taxon:3708"
/clone="EL2449"
/dev_stage="Mid to late embryos (4-6 mg)"
/lab_host="E. coli ElectroMAX DH5 alpha-e (Invitrogen)"
/clone_lib="Brassica embryo library (EL)"
/note="Organ: Embryos without seed coat; Vector: pSPORT1 (modified: GCGGCCGCCc*GACTAGTGAGCTC*cggacgcgtggGTCGAC); Site_1: NotI; Site_2: SalI; Seeds were collected by Dr. Francois Ouellet when they were still very green (mid to large stage, cotyledons were formed). The seed coats were removed and the remaining tissue was used for cDNA library construction. mRNA was poly-A primed using SuperScript Plasmid System cDNA Synthesis and Cloning kit (Invitrogen) After initial screening, the most abundant redundant clones were screened out using 22 oligos designed to match napins (including albumins), cruciferins, oleosins, trypsin inhibitor 2, cytosolic GAPDH, cyclophilins, HSP70, desaturase, and CAB (LHCP)."

ORIGIN

Alignment Scores:

Pred. No.:	2.98e+03	Length:	514
Score:	42.00	Matches:	6
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	7	Gaps:	0

US-09-858-332G-28 (1-6) x CN827673 (1-514)

```
Qy          1 HisAsnHisAsnHisAsn 6
              ||||||||||||||||||
Db        135 CATAACCATAACCATAAC 118
```

Search completed: November 28, 2005, 12:35:09
Job time : 2675 secs